



GÖTEBORGS UNIVERSITET

Big Data

En undersökande studie om hur Big Data uppfattas hos IT organisationer i Sverige

Big Data

An exploratory study on how Big Data is perceived by IT organizations in Sweden

Mathias Jansson Pessi
Sebastian Mueller Moracanin

Examensarbete/Kandidatuppsats i systemvetenskap

Rapport nr. 2014:053
ISSN: 1651-4769

Abstrakt

Datamängden ökar kontinuerligt inom organisationer. Data kommer även i större mängd, får kortare livslängd, blir mer varierad och innehåller olika grader av osäkerhet. Tillsammans ingår dessa dimensioner i termen Big Data. Det finns ingen allmänt accepterad definitionen av Big Data eller hur Big Data skall användas. Tillämpningar som visar sig ingå i Big Data konceptet börjar ofta som satsningar i en organisation för att komma till nya insikter eller skapa nya perspektiv. Syftet med studien är att jämföra olika teoretiska beskrivningar av Big Data och relaterar de gentemot hur svenska organisationer tillämpar IT lösningar vilka kan definieras som Big Data. Studien omfattar fyra intervjuer med olika aktörer som arbetar med Big Data tillämpningar. Resultatet av analysen visar stora skillnader mellan hur organisationer uppfattar Big Data, vad det kan användas till och vilka utmaningar och risker som finns vid dess tillämpning. Studiens bidrag till ämnesområdet informatik är ökad kunskap om fenomenet Big Data. Vidare har vi även kommit fram till att Big Data kan användas som ett nytt komplement av datakällor i beslutsstödssystem. Uppsatsen är skriven på svenska.

Nyckelord: big data, business intelligence, beslutsstöd, riskhantering, kunskapsöverföring

Abstract

Data sets are continuously increasing within organizations. Data sets will also be in larger amounts, have shorter lifetime, is becoming more varied and holds various degrees of uncertainty. Together, these dimensions are included in the term Big Data. There is no universally accepted definition of Big Data and how it should be used. Applications that turn out to be part of a Big Data concept often begin as investments in an organization in order to come to new insights or create new business perspectives. The purpose of this study is to compare different theoretical descriptions of Big Data and relate them to how Swedish organizations apply IT solutions which can be defined as Big Data. The study includes four interviews with various actors involved in Big Data applications. The results of the analysis show significant differences between how organizations perceive Big Data, what it can be used for and what challenges and risks involved in its implementation. The study's contribution to the field of Informatics is a better understanding of the phenomenon of Big Data. Furthermore, we have also come to the conclusion that Big Data can be used as a new complement of data sources in decision support systems. This thesis has been written in Swedish.

Keywords: big data, business intelligence, decision support system, risk management, knowledge

Tack

Vi vill tacka våra respondenter som tagit sig tid och ställt upp på intervjuer vilka gav oss ovärderligt material till uppsatsen. Tack även till vår handledare Lennart Petersson som har väglett oss under arbetets intensivaste stunder. Slutligen även ett tack till examinatorn Dina Koutsikouri för bra konstruktiv kritik.

Innehållsförteckning

1.Introduktion.....	1
1.1 Bakgrund.....	1
1.2 Tidigare forskning.....	3
1.3 Problem.....	3
1.4 Syfte & frågeställning.....	3
1.5 Definition och avgränsning.....	4
1.6 Målgrupp.....	4
1.7 Teoretisk referensram och undersökningens upplägg.....	4
2.Teori.....	5
2.1.Introduktion till Big Data.....	5
2.2.Big Data: Dimensioner.....	7
2.3.Big Data: Datastruktur.....	10
2.4.Relaterade ämnesområden.....	12
2.4.1.Data, information och kunskap.....	13
2.4.2.Beslutsfattning och IT-stöd.....	13
2.4.3.Business intelligence.....	14
2.4.4.Generationer av beslutsstöd.....	15
2.4.5.Business intelligence kopplat till Big Data.....	15
2.4.6.Data Warehouse och Data Mining.....	16
2.4.7.Open data.....	17
2.4.8.Social Media.....	17
2.4.9.Internet of Things.....	18
2.4.10.Quantified Self.....	18
2.5.Tillämpningsområden.....	18
2.6.Nytta och effekter.....	24
3.Metodval.....	32
3.1.Vetenskaplig ansats.....	32
3.2.Datainsamling.....	33
3.2.1.Litteraturstudier.....	33
3.2.2.Intervjuer.....	34
3.2.3.Tematisk analys.....	35
3.2.4.Urval av informanter.....	36
3.2.5.Genomförande.....	36
3.3.Analys av kvalitativ data.....	36
3.4.Källkritik.....	37
4.Resultatdiskussion.....	38
4.1.Presentation av respondenter.....	38
4.2.Definition och syn på Big Data.....	39
4.3.Nytta och effekter.....	44
4.4.Risker, barriärer och utmaningar.....	49
5. Analys.....	53
5.1.Diskussion.....	53
5.1.1.Vad är Big Data och hur förhåller det sig till andra liknande begrepp?.....	53
5.1.2.Vilka tillämpningsområden anses finnas för Big Data?.....	54
5.1.3.Hur beskrivs nyttan och potentiella effekter av Big Data?.....	54

5.1.4. Vilka utmaningar finns när det gäller Big Data.....	55
5.1.5. Vilka risker och barriärer upplevs med Big Data?.....	55
6. Slutsats.....	56
6.1. Förslag till fortsatt forskning.....	56
7. Källor.....	57

Bilaga 1 - Intervjumall v. 1.0

Bilaga 2 - Intervjumall v. 1.1

1.Introduktion

Detta kapitel introducerar bakgrunden till varför vårt arbete är intressant. Vi beskriver den tidigare forskningen inom ämnesområdet och relaterar det med uppsatsens syfte och relevans. Kapitlet beskriver även undersökningsområdet med relevans till vårt syfte och vi definierar en del begrepp samt hur vi har valt att avgränsa oss. Målgruppen för vårt arbete definieras och slutligen beskriver vi upplägget för vår undersökning.

1.1 Bakgrund

Att datamängder kontinuerligt ökar skriver nog de flesta under på, det kan sägas vara en vedertagen sanning. På de senaste decenniet har dock datamängden ökat lavinartat vilket har medfört nya utmaningar att hantera och bearbeta data. Svårigheten att hantera och bearbeta data med dagens verktyg har fått termen Big Data och är ett område för kontinuerlig och intensiv forskning (Bhatia, 2013). Svårigheten att hantera data bekräftas av Gobble (Gobble, 2013) där hon anser att organisationer och samhället drunknar i data. En större acceptans av sociala medier och utbredningen av sensorer medför enorma mängder data i varierande strukturer (Ching-Hsien, 2014). Vidare bekräftar även Tien (Tien, 2013) att Big Data är en term som tillämpas på datamängder som är större än de verktyg som finns idag för att hantera datamängden. Andra, som till exempel Pflugfelder (Pflugfelder, 2013), menar att Big Data är svårt att definiera utan en kontext. Definitionen av Big Data kan göras först när kontexten där data skall användas är känd. Även om kontexten inte alltid är känd, eller är diffus, så finns så anser Cumbley (Cumbley, 2013) att det finns en drivkraft från organisationer att analysera och exploatera informationen som finns i Big Data. Mycket av den forskning som finns inom Big Data idag fokuserar på tekniska aspekter. Ett tekniskt fokus gör att ett Big Data projekt anses tillhöra IT verksamheten men Frankel (Frankel, 2012) anser att man istället borde utgå från verksamhetens behov vid tillämpning av Big Data relaterade projekt.

Det råder således mycket spridda definitioner av Big Data som begrepp, vad det innebär samt hur det skall användas. Något mindre omstritt är att intresset för Big Data har successivt ökat sedan 2010 för att explodera runt 2012. Det generella intresset för Big Data över tid från 2004 till april 2014 beskrivs i figur 1.

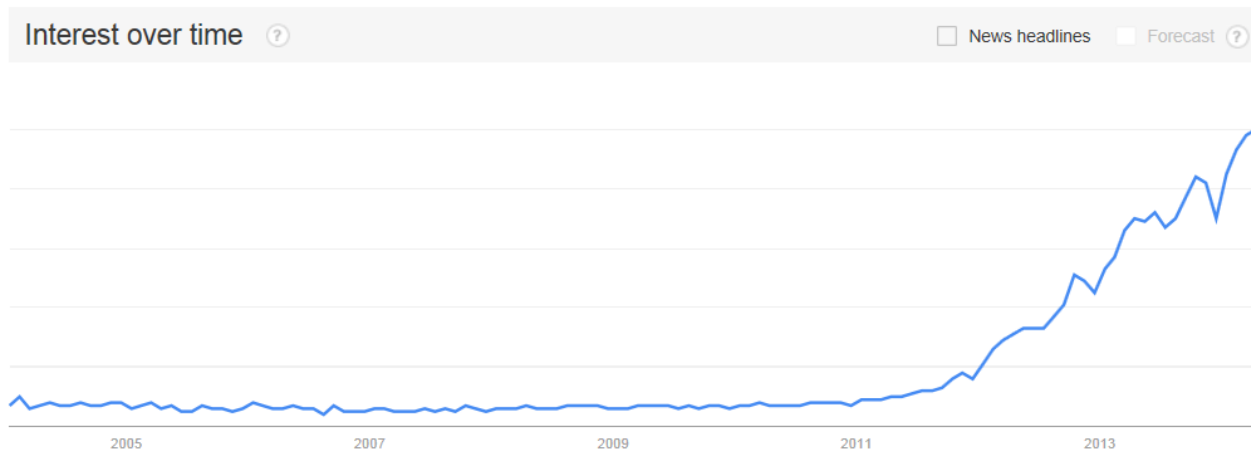


Fig. 1 Intresse för Big Data över tid, källa Google Trends

Intresset för Big Data har ökat betydligt relativt mellan årtalen 2010, 2011, 2012, 2013 till 2014 och beskrivs av Google Trends som en stadig uppgång i figur 2. Bilden (fig. 2) visar staplar till vänster som även visualiseras med grafer till höger. Intresset för Big Data för 2010 visas av den lägsta stapeln varav den för 2014 (fram till April) är den största. Trenden för sökningar efter termen Big Data ökar således från år till år.

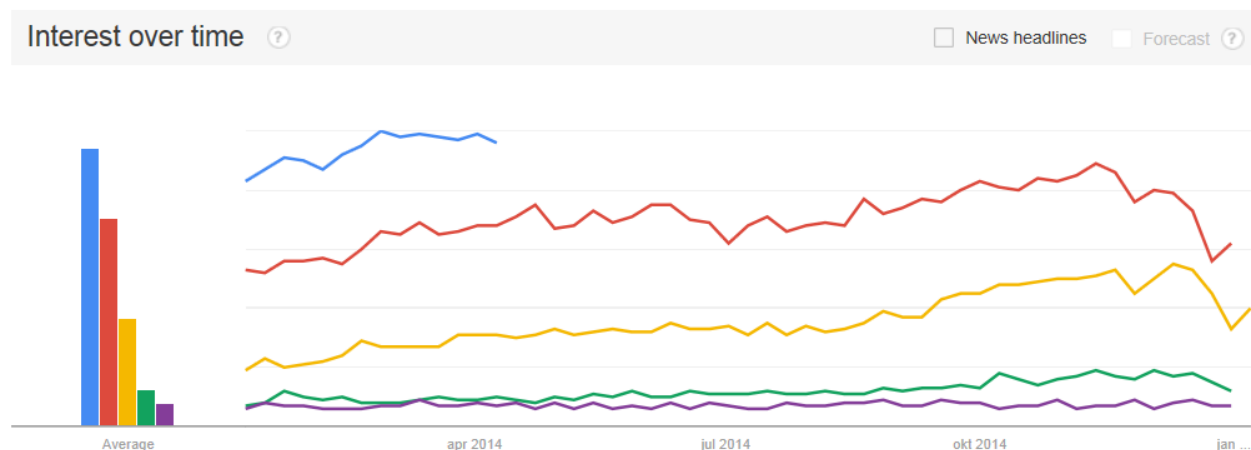


Fig. 2 Intresse för Big Data som sökterm från 2010 fram till April 2014, källa Google Trends

1.2 Tidigare forskning

Forskningen inom Big Data har främst varit inom tekniska lösningar samt hur data skall lagras. Resultatet har varit att det finns lösningar som begränsas av hårdvara istället för mjukvara som t.ex. NoSQL, MapR och Hadoop. Dessa är relativt väl utvecklade men det återstår en hel del utveckling på en mer övergripande nivå.

1.3 Problem

Idag finns det ingen erkänd standarddefinition av begreppet Big Data enligt Chen (Chen, et. al., 2014). Bristen på definition gör det intressant för oss att undersöka vad begreppet innebär. Dock är innebörden endast en del av vårt problem då vi även vill veta om Big Data kan vara värdefullt och om det finns några risker involverade vid tillämpning. Baserat på vårt problemområde riktar vi in oss på en helhetsbild och övergripande förståelse för begreppet och tillhörande ämnesområden som business intelligence (beslutsstöd).

1.4 Syfte & frågeställning

Syftet med studien är att skapa en ökad kunskap och förståelse för fenomenet Big Data, dess tillämpningsområden, nytta samt de utmaningar och risker som förknippas med fenomenet. Med utgångspunkt från studiens syfte blir uppsatsens huvudfråga explorativ till sin karaktär.

- Vad är Big Data och vilka tillämpningsområden, nytta, utmaningar samt risker är relaterade till Big Data?

1.5 Definition och avgränsning

Arbetet har avgränsats till att belysa Big Data som närliggande ämnesområde som business intelligence, kunskapsgenerering samt riskhantering. Tekniska aspekter av Big Data, såsom lagringsstrukturer, hårdvara och allehanda statistiska, matematiska algoritmer kommer ej att behandlas. Ord som används ofta och har en nyckelroll i uppsatsen väljer vi att inte översätta då mycket handlar om att undersöka själva begreppet och inte betydelsen av ordet. Big Data är ett sådant ord. De ord som vi faktiskt har översatt, där det är lämpligt, är ”business intelligence” som benämns ”beslutsstöd”, ”risk management” översätter vi till ”riskhantering” samt ”knowledge management” översätter vi till ”kunskapshantering”.

1.6 Målgrupp

Uppsatsen är främst ämnad för systemvetare och datavetare inom akademien. Systemvetenskapen kan få nytta av rapporten genom att se hur begreppet definieras i näringslivet samt hur begreppet relaterar till områden som beslutsstöd och riskhantering. Datavetenskapen kan få en inblick i hur Big Data tolkas och används idag. Vetenskapen kan således utnyttjas till att bygga nya tekniska implementationer eller utveckla befintliga.

1.7 Teoretisk referensram och undersökningens upplägg

Den teoretiska referensramen syftar i första hand på att skapa en djupare förståelse av begreppet Big Data som stöd för utformning av den empiriska studiens struktur och frågor. Den teoretiska referensramen är indelad i två logiska delar. Den första delen är konceptuell och belyser hur Big Data definieras samt vilka dimensioner, datatyper och datastrukturer som tas upp av litteraturen. Vidare beskrivs hur begreppet förhåller sig till andra närbesläktade begrepp såsom business intelligence, data warehouse samt andra aktuella begrepp och fenomen. Denna första del är relaterad till uppsatsens första delfråga och utgörs av delavsnitt 2 - 2.3. Den andra delen är relaterad till uppsatsens övriga 4 delfrågor (fråga 2-5) och belyser tillämpningsområden, nytta och effekter, utmaningar samt risker och barriärer. Den andra delen utgörs av delavsnitt 2.4 – 2.5.

2. Teori

Detta avsnitt redogör en övergripande bild om Big Data och relaterade ämnesområden. Först introducerar vi vad andra författare skriver om Big Data samt dess beståndsdelar såsom dimensioner och datastrukturer. Relaterade ämnesområden beskrivs och varför de är relevanta för Big Data. Vidare beskriver vi hur kunskap skapas i en organisation. Därefter beskriver vi hur kunskapen tillämpas med hjälp av beslutsstödssystem understött av Big Data. Avslutningsvis beskriver vi riskhantering och varför det är intressant inom Big Data sammanhang.

2.1. Introduktion till Big Data

Det finns ingen allmänt accepterad definition av Big Data då begreppet tycks vara relativt till hur avancerad teknologin i samhället är. Ross (Ross, 2013) anser att det troligtvis inte kommer att finnas en fullständigt accepterad definition av Big Data. Han menar dock att vi kanske inte skall vara för bekymrade över det utan försöka förstå och utveckla tillämpningen av det och så får definitionen utvecklas med tillämpningen (Ross, 2013). Ett vanligt sätt att definiera Big Data är att relatera det till datamängder och databassystem. Andemeningen är att Big Data refererar till datamängder som är så stora att de överskrider traditionella databassystems förmåga/kapacitet att hantera dessa. Emellertid har flera försök till att definiera begreppet gjorts och nedan presenteras och diskuteras ett urval av dessa.

"Big Data refers to datasets whose size is beyond the ability of typical database software to capture, store, manage, and analyze."

- McKinsey Global Institute (Manyika et. al., 2011, sid. 1)

"Big Data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures."

- O'Reilly Radar (Dumbill, 2012, sid. 1)

En annan, liknande typ av definition fokuserar på information som inte kan bearbetas eller analyseras med hjälp av traditionella processer/processorer och verktyg.

”Big Data applies to information that can't be processed or analyzed using traditional processes or tools.”

- IBM Corporation (Zikopoulos, et. al., 2012a, sid. 3)

En tredje typ av definition fokuserar på de utmaningar som organisationer står inför när de skall hantera mycket snabbt växande mängd datakällor och information.

”Big Data is about the growing challenge that organizations face as they deal with large and fast-growing sources of data or information that also present a complex range of analysis and use problems.”

- IDC (Villars, R., et. al., 2011, sid. 2)

Någonting som verkar gemensamt för definitionerna är att dagens databaser och analyskapacitet inte klarar av att hantera dessa enorma mängder data och benämns därför som Big Data. Russom (Russom, 2011) nämner dock att fokuset tidigare har legat på själva datavolymen och att Big Data bäst förklaras genom ”V:na” vilka förklaras närmre i kapitel 2.2.

Enligt Manyika et. al. (Manyika et. al., 2011) är definitionen av Big Data avsiktligt subjektiv då det helt enkelt inte går att bestämma en gräns för storleken på datamängden utan att den teknologiska utvecklingen hela tiden går framåt och där även uppfattningen om vad som anses för stort för att kunna vara hanterbart ingår.

Definitionen skiljer sig även mellan olika branscher då de hanterar olika typer och har olika krav på data. Någonting som verkar gemensamt för definitionerna är att det handlar om stora, snabbt växande och delvis nya datamängder, datatyper, dataflöden och datakällor. Vidare att dagens databassystem och processorskapacitet inte klarar av att hantera dessa större mängder data och därför benämns datamängderna som Big Data. Utvecklingen inom digitala sensorer, kommunikation, processorer, lagring och datafångst gör det möjligt att generera denna stora mängd data (Rendal et. al., 2008). Man förväntar sig en utveckling av teknologi och nya metoder för att kunna analysera och använda Big Data (Davenport, 2012). Ovan definitioner är formulerade av analysföretag eller teknikleverantörer, vilket kanske ger en speciell vinkling av begreppet. En mer omfattande definition, som är publicerad i akademiskt sammanhang, ges av Boyd och Crawford i (Boyd & Crawford, 2012, sid. 663).

”We define Big Data as a cultural, technological, and scholarly phenomenon that rest on the interplay of:

*(1) **Technology**: maximizing computer power and algorithmic accuracy to gather, analyze, link, and compare large data sets.*

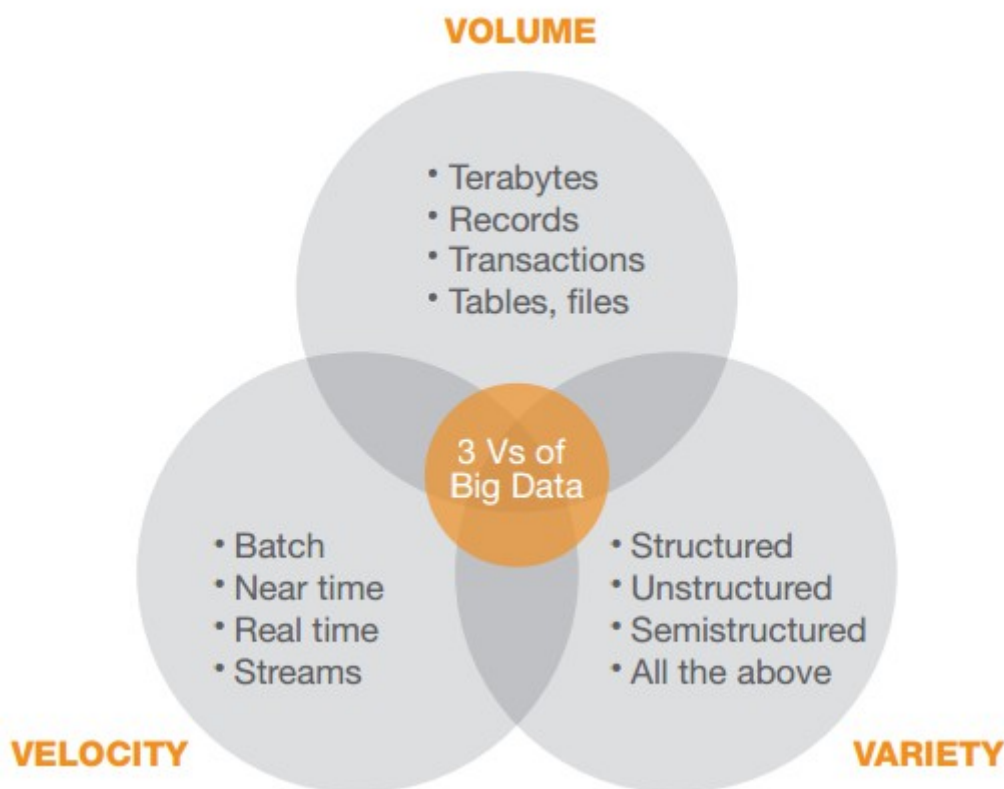
*(2) **Analysis**: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims.*

*(3) **Mythology**: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy.”*

Det ligger utanför denna uppsats ram att generera en egen definition av begreppet Big Data, utan snarare att belysa skillnader och hur begreppet uppfattas i IT organisationer i Sverige. Syftet med det här avsnittet har varit att skapa en förståelse för hur begreppet kan definieras av olika aktörer och att försöka tolka innebörden av dessa definitioner.

2.2. Big Data: Dimensioner

Definitioner av begrepp kan ofta bli väldigt allmänna och generella. När det gäller nya fenomen, kan det vara svårt att skapa en allmänt accepterad definition. Ett sätt att få en fördjupad förståelse är att studera vilka attribut, karakteristika eller dimensioner som förknippas med begreppet. Stora mängder av varierande data kan initialt ge intrycket av att vara svåra att överblicka och få grepp om. Komplexiteten i datamängden kan reduceras genom att kategorisera data i dimensioner (Dumbill, 2012; Ohlhorst, 2013). Flera verksamma personer som jobbar inom fält som kan anses vara Big Data (Bhatia, 2013; Dumbill, 2012; Ohlhorst, 2013; Russom, 2011) nämner att Big Data ofta förknippas med stora datavolymer men att det är endast en av flera dimensioner vilka kännetecknar Big Data. De menar att även om fokus tidigare har legat på själva datavolymer, så kan Big Data bäst förklaras genom att komplettera med ytterligare två dimensioner, vilket sammantaget utgör de tre ”V:na”: Volym, Velocitet, Variation (eng. Volume, Velocity and Variety). Dimensionerna illustreras i figur 3.



Figur 3: De 3 V:na inom Big Data (Russom, 2011)

Volym - datamängd

Ohanterligt stora mängder data är ett kännetecken för Big Data. Stora verksamheter kan hantera datamängder i volymer om terabytes eller t.o.m. petabytes av information medans mindre verksamheter inte nödvändigtvis behöver hantera lika stora mängder data. Rimligtvis anpassar en organisation flödet av data relativt till sin storlek (Dumbill, 2012; Ohlhorst, 2013).

Volym är utan tvekan den primära dimensionen av Big Data och de flesta kvantifierar det i terabytes (TB) eller i vissa fall till och med petabytes (PB) och zettabytes (ZB) (Zaslavsky, et. al., 2012). Volym går även att kvantifiera i form av att räkna antal transaktioner, filer, tabeller och poster (eng. records). Vissa organisationer väljer att kvantifiera datavolym ur ett tidsperspektiv, exempelvis i USA där många organisationer väljer att spara sju års data då det är preskriptionstiden för att kunna vidta rättsliga åtgärder kring en händelse. Omfattningen av Big Data kan påverka datakvantiteten då det exempelvis kan skilja sig stort mellan insamlad data för traditionell data warehousing och insamlad data för analys. (Russom, 2011)

Variation – datastrukturer och källor

Olika typer av data såsom strukturerad (t.ex. databaser i tabellform) och ostrukturerad (t.ex. bilder) blandas i de stora mängder data som utgör Big Data. Semistrukturerad data har delvis en struktur men behöver ett sammanhang för att vara intressant (Dumbill, 2012; Ohlhorst, 2013). En av utmaningarna med Big Data är att data kommer från fler källor än någonsin. Nya källor är t.ex. sociala medier, mobila plattformar och sensorer (Internet of Things). Organisationer har samlat på sig stora mängder data länge men det är först nu som datan börjar analyseras istället för att enbart samlas in och lagras. Big Data är därför inte något helt nytt men att kunna göra effektiva analyser av den enorma mängden data är däremot det. Traditionell strukturerad data får då sällskap av semi- och ostrukturerad data från de nytilkomna källorna vilka behöver mer avancerade analyser för att utvinna kunskap ifrån (Russom, 2011). En mer utvecklad kategorisering av variation ges i (Collaborative, 2012) i termer av: ”Big Interactions”, ”Big Transactions”, ”Big Processing” och ”Big Analytics”.

Velocitet – hastighet av datagenerering

Med velocitet menas hastigheten av datagenerering och frekvensen av datahämtning. Zaslavsky et. al. (Zaslavsky, et. al., 2012) identifierar tre huvudkategorier: Tillfälliga (eng. occasional), Frekventa (eng. Frequent) och realtid (eng. Real-Time). Exempel på datagenerering kan vara strömmande data från termometrar vilka känner av temperaturförändringar, mikrofoner som lyssnar efter rörelse eller videokameror vilka söker igenom folksamling efter specifika ansikten. En särskild form av data är data som genereras av maskiner som ”pratar” med andra maskiner (Collaborative, 2012). Att samla in realtidsdata är inget nytt men att kunna analysera denna stora mängd data i realtid kan vara en utmaning, speciellt då det eftersträvas att data skall analyseras, tolkas och generera en reaktion i realtid. Ett bra exempel på användning av realtidsdata är e-handelshemsidor där data kring besökare analyseras i realtid för att exempelvis kunna föreslå produkter som passar besökaren baserat på tidigare köp och produkter som har kollats på. (Russom, 2011)

Ytterligare dimensioner: Sanningshalt, Värde och Viabilitet

Volym, variation och velocitet är tre dimensioner som ofta återkommer inom litteraturen kring Big Data (Russom, 2011; Dumbill, 2012). Fler aktörer som t.ex. Ohlhorst, 2013 och Corrigan et al (Corrigan et. al., 2012) har valt att lägga till en fjärde dimension av Big Data. Denna fjärde dimension benämns **sanningshalt (eng. veracity)**. Med sanningshalt i Big Data menas kvalitet och tillförlitlighet. Den enorma mängden data vilken samlas in kan innehålla otillförlitlig data (eng. data noise) och exempel på detta kan vara spam vilket lätt kan leda till statistiska fel och feltolkning av data. Bland annat Corrigan och Ohlhorst (Corrigan et. al., 2012; Ohlhorst, 2013) menar därför att det är viktigt att försäkra sig om en hög datakvalitet och trovärdiga källor för att värdet av data skall bli så högt att organisationer både kan, men framförallt vågar fatta beslut med data som underlag. Corrigan et. al. är ännu mera specifika och menar att för säkerställande av sanningshalten i data behövs verktyg för att kunna göra data mer pålitlig genom att reducera data noise (Corrigan et. al., 2012). **Värde (eng. Value)** anges av Ross (Ross, 2013) som en viktig dimension. Han menar att en kritisk nyckelfaktor för Big Data är att organisationer måste kunna se stora värden direkt från deras data. Ross tar även upp en sjätte dimension, **viabilitet (eng. viability)** som innebär förmågan att snabbt identifiera data som innehåller användbar information (Ross, 2013).

2.3. Big Data: Datastruktur

I föregående underkapitel 2.2 nämner vi att en dimension av Big Data är variation. Dimensionen består av ett antal olika datastrukturer som behöver hanteras och analyseras. Variationerna är strukturerad, semi och ostrukturerad data (Ohlhorst, 2013; Russom, 2011). Datastrukturen är en viktig faktor som bestämmer hur mycket arbete som måste läggas på att utvinna information ur datamängden. Ross (Ross, 2013) skiljer även mellan extern data och intern data. Nedan ges en närmare beskrivning av de vanligaste datastrukturerna baserat på Russom (Russom, 2011).

Strukturerad

Data i strukturerad form kan vara behandlad och ordnad i tabeller i en annan databas som bara kopieras in i ett datalager (eng. data warehouse). Strukturerad data är den mest förekommande datatypen i dagsläget. (Russom, 2011) Text och rådata som finns i ett strukturerat format såsom SQL databaser, textfiler och loggar. Denna datastruktur ingår i den första generationen av stödsystem (BI 1.0) (Chen et. al., 2012). Realtidsdata kan innefatta data från sensorer av olika slag. Det kan också handla om data som genereras när maskiner ”pratar” med varandra. Spatial data från bl.a. GPS’er. Anger spatiala positioner i form av longitud/latitud koordinater, etc.

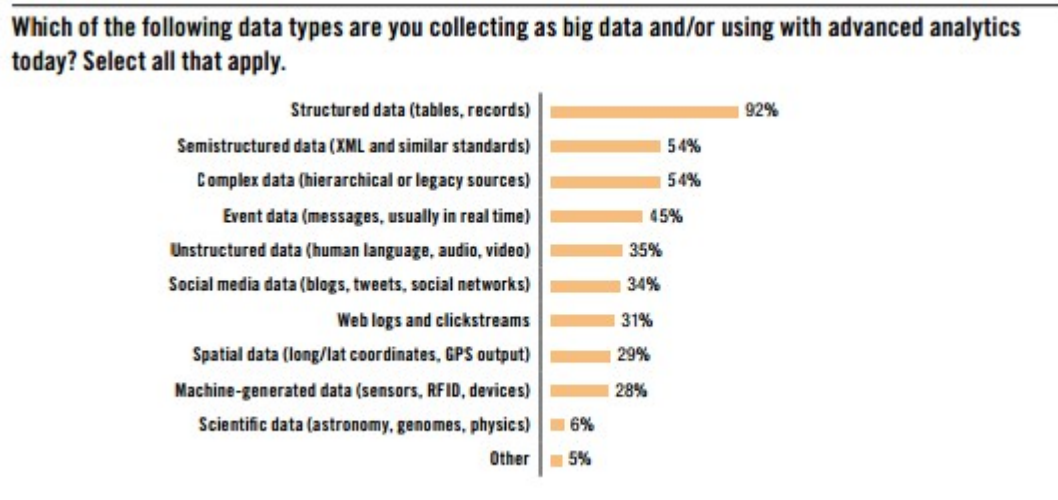
Semistrukturerad

Semistrukturerad data behöver sättas i någon form av kontext för att kunna tolkas. Exempel på semistrukturerad data är inlägg i sociala medier och textkonversationer som har ett större sammanhang (som t.ex. bilforum). Särskilda verktyg behövs för att extrahera informationen och placera den i rätt kontext. Oftast finns metadata tillgänglig vilket kan förenkla att sätta datamängden i en kontext. Ett sätt att kontextualisera semistrukturerad data är med hjälp av märkspråket XML. Semistrukturerad data är en datastruktur som är under stark frammarsch inom organisationer och blir allt vanligare att utnyttjas som informationskällor. (Russom, 2011) Både semistrukturerad och ostrukturerad data ingår i den andra generationen av stödsystem (BI 2.0). Web data innefattar data från sociala medier som exempelvis bloggar, tweets och sociala nätverk men även webb loggar och klickströmmar (eng. Clickstreams).

Ostrukturerad

Ostrukturerad data kan vara information vilken finns i bilder, filmer, tal och även texter som kräver avancerade analytiska verktyg för att kunna analyseras. Information som finns lagrad i bilder och filmer kan inte avläsas i sin digitala, obehandlade form. Datamängden behöver någon form av behandling för att kunna utläsas och tolkas (Russom, 2011). En programvara kan behandla data och automatiskt identifiera ansikten och personer i digitala bilder. Programvaran kan sedan extrahera informationen och jämföra den mot externa datakällor för att på så sätt få fram t.ex. identitet, relationer, efterlysningar, registreringar etc. (Ohlhorst, 2013).

Enligt en undersökning av Russom (Russom, 2011) så dominerar fortfarande strukturerad data, följt av semi-strukturerad data och komplex data (se figur 4 nedan).



Figur 4: De vanligaste datatyperna som samlas för Big Data. (Russom, 2011)

2.4. Relaterade ämnesområden

Två begrepp som är nära relaterade till Big Data är business intelligence och Data Analytics. I en introduktion till ett specialnummer om business intelligence forskning har Chen et.al. (Chen, et. al., 2014) rubriken:

”Business Intelligence and Analytics: From Big Data to Big Impact”

Med det menar de att business intelligence och Analys (BI&A) är de tekniker, teknologier, system, praktik, metodiker och applikationer som används för att analysera kritisk affärsdata. Detta i syfte att bättre förstå affärsverksamheten och marknaden samt att i tid fatta beslut. BI&A är sålunda fokuserad på att analysera och tolka Big Data i syfte att kunna agera klokt. Begreppet ”Big Data Analytics” är nära kopplad till business intelligence enligt Chen et. al. (Chen, et. al., 2014). Det är ett område som utvecklas i takt med att enormt stora mängder data och dataströmmar görs tillgängliga. Russom beskriver Big Data Analytics på följande sätt:

”Big Data analytics is the application of advanced analytics techniques to very Big Data sets” (Russom, 2011, sid. 4)

I den följande texten beskriver vi business intelligence närmare och hur det utvecklats till det som Chen et. al. (Chen, et. al., 2014) kallar för ”Business Intelligence and Analytics”.

2.4.1. Data, information och kunskap

Ny data genereras hela tiden i många sammanhang. Loggar, bilder, text etc. genererar data i organisationer och som öppna data på Internet vilket finns i stora mängder med varierande grad av tillgänglighet. Data som lyfts fram och behandlas blir till information om den sätts in i ett sammanhang. Hur det sker är beroende på vilket syfte som finns med databehandlingen (Ohlhorst, 2013). Att leta igenom olika typer av data kräver särskilda verktyg för att integrera med ett beslutsstöds system (Turban, et. al., 2011). Till skillnad från semistrukturerad data så är ostrukturerad data ännu svårare att hitta nytta i vilket kräver ännu mer avancerade verktyg för att kunna kategorisera och hitta nyttig data (Ohlhorst, 2013). För att kunna få verksamhetsnytta, eller ett värde, behöver data omarbetas till information som är intressant för sammanhanget.

En process som kan användas med Big Data, för att hitta nytta, är ETL (Extract, Transform and Load) vilken bygger upp ett datalager där en analys av data kan tillämpas. Termen innebär en process där data inhämtas (Extract), sällas och behandlas (Transform) och slutligen infogas i en databas (Load). ETL processer kan användas för att bygga strukturerade databaser och kan användas till att bryta ner affärsprocesser till ett ETL lager (process layer) (Turban, et. al., 2011). ETL processer används ofta till att infoga information till ett datalager (data warehousing, DW) vilket kan vara en understödjande källa till verksamhetens business intelligence.

2.4.2. Beslutsfattning och IT-stöd

När en ETL process har fört över data till ett datalager inom en business intelligence domän i en organisation kan en analytisk process utföras. Analysen som bör ske i nästan realtid, utförs och ett beslut kan fattas på hur systemet skall gå vidare. Kopplingar mellan verksamhetsprocesser som tidigare inte var kända kan upptäcka och leda till nya idéer och utvecklingsmöjligheter (Dijcks, 2012). Det är inte rimligt att göra allt och kunna allt, speciellt om allt skall göras samtidigt. Att hitta vad som är intressant i Big Data anser Ohlhorst (Ohlhorst, 2013) vara som att hitta en nål i en höstack. Big Data projekt som är direkt relaterade till affärsnytta kan bidra med svar på frågor knutna till verksamhetsprocesser. I många fall är inte bakomliggande logik av verksamhetsprocesser särskilt enkel utan har en betydligt komplexare struktur i form av t.ex. sociala, tekniska och strukturella hinder (Ohlhorst, 2013; Turban, et. al., 2011). Ohlhorst (Ohlhorst, 2013) menar att en generalisering av data kan potentiellt vara förenat med missade affärstillfällen. Ett rimligt antagande kan då vara om det är lönsamt att ignorera vissa missade affärstillfällen.

2.4.3. Business intelligence

Organisationer befinner sig i en ständigt föränderlig omvärld. Förändringar i omvärlden kan härstamma från olika omständigheter, exempelvis kunders förväntningar, konsumtionsmönster samt konkurrenters verksamhet. För att organisationer skall kunna verka effektivt i denna turbulenta miljö krävs verktyg och metoder för att agera snabbt och agilt. Omvärldens förändringar bör bemötas på ett objektivt sätt genom att samla in data, information och kunskap som kan understödja datoriserade beslutsstödssystem (Turban, et. al., 2011). Beslutsstöd som fenomen sammanfattas väl i följande stycke:

“[Business intelligence] deals with a collection of computer technologies that support managerial decision making by providing information on internal and external aspect of operations. These technologies have had a profound impact on corporate strategy, performance, and competitiveness. These technologies are collectively known as business intelligence.” (Turban, et. al., 2011, sid. 21)

Således ger beslutsstöd affärsutvecklare verktygen att bedriva affärsanalyser baserat på data. Data som används i affärsanalyser kan vara historisk eller i realtid vilket ger affärsutvecklare möjligheten att fatta informerade beslut istället för kvalificerade gissningar.

Ett företag kan beskrivas som ett antal processer. Beslutsstöd tillämpas vanligtvis inom en process eller knyter samman information från flera processer, vilket kan möjliggöra ny kunskap. Det är vanligt att analysera transaktioner genom att mäta försäljning, lagersaldo och leveranstider (Turban, et. al., 2011). Analyser som genomförs inom ramen för beslutsstöd kan användas i många olika scenarion. Ett system som automatiskt uppdaterar lagersaldo vid försäljning ger möjligheten att göra en analys på lagersaldo och rekommendera en åtgärd, eller automatiskt skicka en beställning till leverantören. En tydlig effekt av en automatisering manifesteras när ett integrerat system på egen hand kan hantera inköp om det känner av att lagersaldot är lågt eller når en viss gräns, vilket kan vara användbart inom t.ex. detaljhandeln. Automatiseringen bidrar till att mindre resurser krävs för att driva och underhålla processer samt att färre fel uppstår genom att minska mänsklig inblandning. Ovanstående beskrivning är en del i ett traditionellt beslutsstöd system som, med en hög grad av automatisering, kan bidra till en stark effektiviseringsgrad av befintliga system. Nyttan som ett beslutsstöd således tillför är bland annat snabba beräkningar och presentation av tillförlitlig data. Väl utvecklade beslutsstöd i en verksamhet kan tillföra nyttiga analysverktyg som underlättar välinformerade beslut.

2.4.4. Generationer av beslutsstöd

På grund av tekniska begränsningar i form av dåligt strukturerad data har det tidigare varit svårt att dra slutsatser baserade på analys av Big Data. I takt med de senaste årens tekniska utveckling av allt kraftfullare hårdvara blir analyser av större datamängder möjliga ur ett tekniskt perspektiv (Ohlhorst, 2013). Kraftfullare hårdvara medför att data från en mängd olika sensorer, datorer, mobiltelefoner och andra källor kan sammanställas och analyseras. Indelningen av datakällor kan kategoriseras till olika generationer av beslutsstöd. Traditionella beslutsstöd är när data finns lokalt producerad i en verksamhet, detta är business intelligence 1.0. Business intelligence 1.0 består ofta av strukturerad data som är betrodd och har hög tillförlitlighet. Nästa generation av beslutsstöd bygger på att sociala medier blir till datakällor. Sociala medier består ofta av semi och ostrukturerad data med text som produceras inom en viss kontext och tillhörande bilder. Om data i sociala medier tas ur sin kontext blir den svår att tolka och kan bli tvetydig. Denna generation av beslutsstöd klassificeras som business intelligence 2.0. Efterföljande generation av beslutsstöd, 3.0, lägger till ytterligare data via mobila enheter och sensorer. Detta kan vara appar i mobiltelefoner, smart hemelektronik och surfplattor som kan kommunicera över t.ex. Internet (Chen et. al., 2012).

2.4.5. Business intelligence kopplat till Big Data

Användandet av Big Data i ett beslutsstöd sammanhang kan komplettera en verksamhet eller process genom att ta med nya, tidigare ej nyttjade, dimensioner. Utgångspunkten för traditionella beslutsstöd är hitta verksamhetsnytta genom att forma och strukturera datamängder baserat på verksamhetsprocesser. Utgångspunkten med Big Data är istället att skapa nya verksamhetsprocesser med hjälp av befintlig data (Ohlhorst, 2013). Data som samlas in kan identifieras och klassificeras genom att titta på sammanhanget där data produceras. Inom ämnet informatik är det vanligt att skapa modeller av verkligheten och infoga dessa modeller som en virtuell representation av verkligheten i en dator (Larman, 2009). Modellen som representerar verkligheten bör vara analyserad och förenklad för att undvika onödigt många variabler i modellen. Trots en förenkling genereras oftast en virtuell modell med analyserbar data. En modell kan analyseras genom att undersöka hur mycket data som finns, när den genererades och vem eller vad som genererade datamängden. En analys kan ge insikter om den information som analyseras, medan en affärsinriktad analys (business analytics) handlar om att analysera om den organisation som utför analysen (Stubbs, 2013). Således kan en informationsmodell även spegla verkligheten genom att ha tillgång till stora mängder data, Big Data.

2.4.6. Data Warehouse och Data Mining

Under framförallt 90-talet var begreppen data warehouse och Data Mining aktuella för att tydliggöra skillnaden mellan operativa informationssystem och system som skapats som stöd för analyser och beslutsfattande. En av de stora förgrundsfigurerna var William Inmon. Han menar att ett data warehouse utgörs av följande (Inmon, 1996):

- Integrerad data
- Detaljerad och summerad data
- Historisk data
- Metadata

Den vanliga synen på data warehouse är att det utgörs av stora strukturerade datamängder hämtade från operativ system, rensade, kvalitetssäkrade, integrerade och framförallt historisk data som integreras, lagras och görs tillgänglig för analys och beslutsfattande i organisationer (Inmon, 1996). Data mining innebär att ”gräva” i den stora mängden data för att göra intressanta upptäckter. Data warehouse är en förutsättning för att det skall kunna göras effektivare. Inmon uttrycker det på följande sätt:

”The data warehouse sets the stage for effective data mining” (Inmon, 1996, sid. 46)

Hur förhåller sig då data warehouse och Data mining till Big Data och BI/Data Analytics. Madden (Madden, 2012) beskriver att det senare handlar om mycket mer än stora databaser. Rubriken på hans artikel är ”From Databases to Big Data”. Stora databaser är endast en del av Big Data, men det täcker inte in alla delar, t.ex. stora dataströmmar och sensor-baserade dataflöden. Begreppet Data Mining är nära förknippad med data warehouse och den datacentriska ansatsen som första generationens business intelligence & Analytics (BI&A 1.0) representerade (Chen et. al., 2012). Chen et al (Chen et. al., 2012) tar upp olika verktyg för att analysera data i data warehouse, såsom rapportfunktioner, statistiska analyser samt data miningverktyg (Chen et. al., 2012). Utvecklingen mot dagens syn på Big Data Analytics beskrivs av Chen et al med begrepp såsom, Web Intelligence, Web analytics, web-mining etc (BI&A 2.0) samt Mobile analytics, location and context-aware analytics techniques (BI&A 3.0). Big Data och BI/Data Analytics inbegriper sålunda nya datatyper, datakällor och dataströmmar utöver data warehouse baserade stora databaser som i huvudsak hanterar strukturerad data.

Cohen m.fl (Cohen, et. al., 2009) presenterar i sin artikel vad de kallar för en uppkommen praktik: ”Magnetic, Agile, Deep (MAD) Data analysis”. De menar att det är en utveckling från traditionell data warehouse och business intelligence. Det nya i deras synsätt är att beakta alla datakällor som kan finnas eller uppstå i en organisation, inte enbart de som är ”tvättade, integrerade” etc. ”Magnetic” symboliserar att man ”drar till sig all sorts data”. ”Agile” innebär att man mycket snabbt skall kunna anpassa och utveckla och analysera data. ”Deep” innebär mycket mer sofistikerade analysverktyg som kan gå betydligt djupare i analysen än tidigare.

2.4.7. Open data

Open Data och Big Data nämns ofta som nära relaterade till varandra. Men det är inte samma sak. Open Data innebär mer ett perspektiv som går ut på att göra data mer användbar, mer demokratisk och tillgänglig för alla. Open Definition definierar Open Data på följande sätt:

“A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.” Opendefinition, 2014, sid. 1)

Mer och mer data görs allmänt tillgängligt, framförallt från myndigheter. Man kan se Open Data som en särskild typ av data källa, som tillsammans med många andra källor tillsammans utgör Big Data.

2.4.8. Social Media

Explosionen av olika typer av Sociala media är en källa till stora mängder av ny data. Chen et. al. menar att sociala media är en del av utvecklingen under Business Intelligens and Analytics 2.0 (BI&A 2.0) (Chen et. al., 2012, sid. 1167):

“The many Web 2.0 applications developed after 2004 have also created an abundance of user-generated content from various online social media such as forums, online groups, web blogs, social networking sites, social multimedia sites (for photos and videos), and even virtual worlds and social games”

Det finns många förväntningar på hur data från sociala media kan användas i många sammanhang. Det kan vara politiska partier som vi känner av hur människor resonerar, företag som vill få feedback på hur kunderna uppfattar företagets produkter och tjänster etc. Sociala media är för många organisationer en ny och stor potentiell datakälla, och inbegrips vanligtvis i begreppet Big Data.

2.4.9. Internet of Things

”Internet of Things” är ett begrepp som refererar till den oerhört stora mängden enheter som kan känna av, kommunicera, beräkna och manövrera saker och ting (Zaslavsky, et. al., 2012). Det är allt ifrån mobiltelefoner, iPads, sensorer, enheter i bilar, kylskåp, RFID utrustade enheter, etc. Dataströmmar från alla dessa enheter är en del av det som kallas Big Data. Det är framförallt denna typ av stora mängder nya dataströmmar som representerar det som Chen et al kallar för Buiness Intelligence and Analytics 3.0 (BI&A 3.0) (Chen et. al., 2012).

2.4.10. Quantified Self

Ett annat begrepp som relateras till Big Data är ”Quantified Self”. Swan definierar det på följande sätt:

”The Quantified Self is any individual engaging in the self-tracking of any kind of biological, physical, behavioral, or environmental information” (Swan, 2013, sid. 85)

Det finns mängder av områden som kan mätas och analyseras t.ex vikt, kaloriintag, kilometertid vid löpning, blodtryck, kroppstemperatur, puls, bara för att bara nämna några. Under de senaste åren har internet formligen exploderat av nya appar och tjänster som kan användas för Quantified Self. En av de mest kända är förmodligen Runkeeper, som många joggare använder för att samla olika typer av data under löpning, såsom kilometertid, totaltid, antal kilometer, höjdskillnader etc. Det finns stora förväntningar på Quantified Self inom särskilt sjukvården, då det ger möjlighet till en stor mängd ny data som kan analyseras för behandling och prevention av sjukdomar. Quantified Self är sålunda en ny källa till en massiv mängd information som kan inbegripas i begreppet Big Data.

2.5. Tillämpningsområden

Efter den konceptuella genomgången av begreppet Big Data och andra relaterade begrepp, fokuseras nästkommande fyra delavsnitt mot en mer tillämpningsorienterad syn på Big Data. Litteraturunderlaget utgörs i huvudsak sekundärdata i form av rapporter från olika forskningsinstitut och s.k. expertorganisationer, men även forskningsartiklar publicerade på akademiska konferenser och tidskrifter.

Det finns givetvis en uppsjö av olika tänkbara tillämpningsområden för Big Data. Vi har valt att fokusera på följande ”branscher”:

- Försäkringsbranschen
- Hälso- och Sjukvård
- Detaljhandeln
- Offentlig förvaltning (Myndigheter)
- Tillverkningsindustrin

Ambitionen i uppsatsarbetet är inte att skapa en heltäckande bild av alla möjliga tillämpningsområden. Vår ambition är i stället att utifrån ovan fem ”branscher” beskriva hur man ser på tillämpningsområden.

Försäkringsbranschen

Försäkringsbranschen är väldigt data-beroende och Big Data är en trend inom branschen. I en rapport från 2013 beskriver Bharal och Halfon (Bharal, 2013) tillämpningsområden för Big Data inom försäkringsbranschen. Det är en bransch som kämpat med hanteringen av data under en lång period, inte minst när det gäller riskhantering. Möjligheterna med Big Data har ökat intresset för hantering och analys av data, då det nu finns än mer källor och flöden tillgängliga. Bharal och Halfon tar upp följande exempel på tillämpningsområden för Big Data (Bharal, 2013):

- Upptäckt/Spårning av bedrägerier (eng. Fraud detection) är ett område, där man försöker skapa ”person-centrerade” ansatser istället för ”anspråks-centrerade” ansatser med hjälp av Big Data. Det innebär datainsamling från många nya källor än tidigare.
- Kundupplevelse är ett annat tillämpningsområde och innebär en liknande perspektivförändring från produkter till kundens upplevelser. Det senare innebär många nya potentiella datakällor och flöden.
- Hantering av försäkringsärenden (eng. Claim management) är ett tredje område, där man ser en trend mot allt mer data och olika format och från många olika källor. Data kan vara strukturerad, semi-strukturerad, ostrukturerad. Det kan vara bilder, videos etc.

- Teckning av försäkringar och då särskilt vid mycket stora åtaganden. Där behövs det analyser av stora datamängder från många olika källor för att kunna göra korrekta bedömningar av olika aspekter, inte minst risker.

Bharal och Halfon menar att Big Data är ett begrepp som är viktigt för försäkringsbranschen att ta till sig och att utveckla kunskapen om hur det kan utvecklas i framtiden.

Hälso- och Sjukvård

Hälso- och Sjukvård är den ”bransch” som oftast nämns när det gäller tillämpningar inom Big Data. I en omfattande studie genomförd av Manyika et. al. (Manyika et. al., 2011) lyfter man fram dagens situation med fyra stora pooler av data inom amerikanska sjukvårdssystemet:

1. Farmaceutisk FoU data
2. Klinisk data
3. Aktivitetsdata och ekonomisk data
4. Data om patientbeteende och patientuppfattningar

En stor utmaning är att integrera dessa olika pooler av data för att utveckla sjukvården. Manyika et. al. (Manyika et. al., 2011) identifierar fem breda områden där det finns potential att förbättra rationalitet och effektivitet genom Big Data. Dessa är:

1. **Klinisk verksamhet**, där de lyfter fram fem delområden där Big Data kan tillämpas. Det första forskningen om jämförbar effektivitet om vilken behandling som fungerar bäst för specifika patienter. Ett annat delområde är kliniska beslutsstöd för att öka kvaliteten och effektiviteten i verksamheten. Det tredje delområdet handlar om att analysera data om medicinska procedurer i syfte att identifiera möjligheter till prestationsförbättringar hos sjukvårds professionen, processer och institutioner. De menar att en transparens av denna typ av data kan underlätta för patienter att välja var man vill bli behandlad. Det fjärde delområdet är fjärrövervakning av patienter, där en hel del data samlas in och kan analyseras på bättre sätt. Det femte delområdet handlar om avancerade analyser som kan kopplas till patientprofiler för att kunna identifiera individer som kan få stor effekt av

proaktiv vård och livsstilsändringar.

2. **Betalning och prissättning** är kanske mer förknippad med den amerikanska sjukvården, men blir allt viktigare även i Sverige. Inte minst gäller det privat kontra offentlig vård. Detta område avser beställarens perspektiv. Manyika et. al. (Manyika et. al., 2011) tar upp automatiserade system som ett delområde för att upptäcka bedrägerier mm. Ganska likt det som gäller försäkringsbranschen (se förra delavsnittet). Det andra delområdet som tas upp är forskning inom hälsoekonomi och utfall samt prestationsbaserade prissättningsmodeller.
3. **Forskning och utveckling** är det tredje området. De delområden som tas upp är: prediktiv modellering för nya mediciner, Statistiska verktyg och algoritmer som förbättrar kliniska provningar, Analys av data från kliniska provningar, personaliserad medicin, och avslutningsvis analys av sjukdomsmönster.
4. **Nya affärsmodeller** är det fjärde tillämpningsområdet, där två potentiella delområden lyfts fram: Aggregering och syntetisering av patientdata i syfte att tillhandahålla data och service till tredje part. Vidare nämns online plattformar och communities, såsom t.ex. PatientsLikeme.com. Dessa nya företeelser kan bli värdefulla källor av data. I Sverige finns kanske mindre förutsättningar för detta då det finns skarpa lagar och regler för hur patientdata kan användas i andra sammanhang än rent kliniska.
5. **Folkhälsa** är det femte området, där Big Data kan förbättra den allmänna bevakningen och agerandet när det gäller folkhälsan. Hälso- och sjukvårdens stora utmaningar är att minska kostnaderna för sjukvården samtidigt som man behöver öka kunskapen och förbättra effektiviteten och effekterna av behandling. Här finns många tillämpningsområden för Big Data. Även den preventiva delen av folkhälsoarbetet är ett viktigt område för Big Data

Offentlig förvaltning (Myndigheter)

Den offentliga förvaltningen är under en ständig press att göra mer med mindre resurser. Kravet på service till medborgarna ökar ständigt. Till skillnad från sjukvården genererar myndigheter mindre multimedia baserad data. Det är mer en ökad mängd textuell och numerisk data (Manyika et. al., 2011). Myndigheter behöver skydda, fördela och analysera både strukturerad och ostrukturerad data för att bättre betjäna medborgarna (O'Brien, 2012). Den offentliga förvaltningen är väldigt stor och diversifierad, vilket gör det svårt att lyfta fram specifika tillämpningsområden i denna uppsats. I litteraturen ges olika exempel på tillämpningar inom Big Data. O'Brien (O'Brien, 2012) ger några exempel från det amerikanska försvaret, energimyndigheten och några andra myndigheter. Manyika et. al. (Manyika et. al., 2011) ger exempel på fem potentiella tillämpningsområden för Big Data:

1. Att skapa transparens genom att en stor mängd data från stora offentliga databaser och andra källor görs tillgängliga. De tar Svenska Skattemyndigheten som exempel, där myndigheten samlar en stor mängd data från olika källor inför medborgarnas deklaration. Medborgaren i sin tur får en för-ifylld deklaration som man bekräftar eller ändrar.
2. Att upptäcka behov, finna variationer, och förbättra myndigheternas prestationer. Big Data kan användas för att jämföra olika enheters prestationer med varandra i syfte att förbättra och utveckla myndighetens verksamhet.
3. Segmentering av populationen för att kunna anpassa service och aktiviteter.
4. Ersätta eller stödja mänskligt beslutsfattande med automatiserade algoritmer.
5. Innovera nya affärsmodeller, produkter och tjänster med hjälp av Big Data. Det kan gälla både inom den offentliga sektorn men även utanför i den privata sektorn.

Den offentliga förvaltningen anses ha många olika tillämpningsområden för Big Data. Inom den offentliga förvaltningen genereras mycket stora mängder data av olika slag. Traditionellt sett har dessa data varit isolerade i transaktionsbaserade system eller myndighetsvisa silo-liknande system. Utvecklingen inom Big Data och Analytics anses skapa nya möjligheter, men också stora utmaningar för den offentliga sidan (O'Brien, 2012).

Detaljhandeln (eng. Retail)

Inom detaljbranschen har IT varit ett medel för att utveckla konkurrenskraft och produktivitet under flera decennier. Många av de klassiska exemplen på framgångsrika data warehouse och beslutsstöds (business intelligence) tillämpningar är från detaljbranschen. Sålunda bör det även finnas potential för tillämpningar inom Big Data, särskilt då mer och mer av handeln sker online via internet och webb, där mängder av data genereras och sparas. Manyika m.fl (Manyika et. al., 2011) menar på att nästa generations detaljist kan följa beteendet hos individuella kunder och modellera deras sannolika beteende i realtid:

”A next-generation retailer will be able to track the behavior of individual customers from Internet click-streams, update their preferences, and model their likely behavior in real time” (Manyika et. al., 2011) sida 7

Manyika et. al. (Manyika et. al., 2011) beskriver följande tillämpningsområden inom detaljbranschen:

Marknadsföring. Korsförsäljning, platsbaserad marknadsföring, analys av kundbeteenden i affären, mikrosegmentering av kunder, annan segmenteringsanalys, utveckla kundens multi-kanal upplevelse.

Försäljning. Optimering av sortiment, prissättnings optimering, optimering av placering och visuell design.

Analys av prestationer och resursbehov vid olika tidpunkter. Exempelvis för att beräkna behov av personal vid såväl hög- som lågsäsong.

Varuanskaffning (supply chain) i form av optimering av distribution och logistik, varulager hantering mm.

Nya affärsmodeller gäller även för detaljbranschen som har möjlighet att samla stora mängder data om kundbeteenden, preferenser etc, direkt från internet. Det kan handla om prisjämförande service för att få kunden att välja rätt detaljist. Det kan också handla om web-baserad marknader, såsom Amazon och eBay.

Detaljbranschen har under de senaste 15-20 åren gått från en fysisk plats-baserad värld till en i högre utsträckning Internetbaserad och platsoberoende värld genom bl.a. den kraftigt ökade e-handeln via Internet. Mängden information som är elektronisk och som potentiellt kan utgöra grund för Big Data är mycket stor, och ökar för var dag.

Tillverkningsindustri

Tillverkningsindustrin genererar stora mängder data från olika delar av verksamheten. Tidigare fysiska produkter digitaliseras mer och mer. Till exempel genererar en modern bil enorma mängder data som kan användas för många olika ändamål. Manyika et. al. (Manyika et. al., 2011) menar att tillverkningsindustrin kan finna tillämpningsområden för Big Data inom hela värdekedjan. Allt från Produktutveckling, Försörjningskedjan, Produktion, Marknadsföring och försäljning till Eftermarknad. Nedan några exempel på tillämpningsområden:

- Hantering av produkters livscykel, produktdesign, öppen innovation.
- Inom Försörjningskedjan (Supply Chain) handlar det om prognostisering av efterfrågan samt försörjningsplanering.
- Inom Produktion talar man om digitala fabriker och sensordrivna aktiviteter.
- Inom marknadsföring, försäljning samt eftermarknad återfinns liknande tillämpningsområden som för detaljister.

Villars et. al. (Villars, R., et. al., 2011) menar att inom tillverkningsindustrin genereras sensor data i en accelererande grad från GPS mottagare, RFID läsare, mobiltelefoner etc. Data som kan användas för att optimera verksamheten och hitta affärsmöjligheter.

Ovan beskrivningar av tillämpningsområden inom olika branscher syftar i första hand till att skapa en bättre bild av inom vilka områden man ser potentialen av Big Data.

2.6. Nyttan och effekter

Big Data och dess användning anses ha en stor och hittills outnyttjad potential för värdeskapande. Många organisationer kan dra nytta av Big Data till att koordinera sina resurser, minska avfall, öka transparens, öka ansvarsskyldighet och möjliggöra upptäckande av nya idéer och insikter. Manyika et. al. (Manyika et. al., 2011) identifierar fem sätt att skapa värde med Big Data vilka ej är specifika för en bransch utan kan appliceras generellt. De nämnda fem generella värdeskapande sätten känns igen från beskrivningen av tillämpningsområden ovan. Dessa fem värdeskapande sätt är:

Skapa transparens.

Det kan skapas ett enormt värde om Big Data bara görs mer lättillgänglig och presenteras i tid för relevanta intressenter. Denna värdeskapande aspekt är ett förkrav för all annan typ av värdeskapande och är det mest omedelbara sättet för organisationer att kunna anamma Big Data och utnyttja dess potential till fullo. Mindre tid krävs då för att ta fram information från andra delar inom organisationen både genom att informationsletandet sker digitalt istället för pappershanterande och även att informationstillgängligheten är förbättrad. Detta gynnar alla inom en organisation, från en chef vilken behöver få fram information om exempelvis prestation i de olika företagsdelarna, till en medarbetare vilken söker relevant information från en annan del av företaget för att kunna jobba effektivt. Manyika et. al. (Manyika et. al., 2011).

Möjliggöra experimentering för att kunna upptäcka behov, hitta avvikelser och öka effektivitet.

De teknologiska möjligheterna för organisationer att samla in data förbättras ständigt och med den ökade digitaliseringen samt insamlingen av data tillkommer även fler källor att hämta data ifrån så som sensorer i enheter. Även konsumenter skapar och delar med sig av data genom olika sociala medier och mycket av dessa data kan insamlas i realtid eller väldigt nära realtid.

Tillgången till dessa data och i vissa fall även att kunna kontrollera omständigheterna kring hur den genereras möjliggör nya sätt att fatta beslut vilka involverar mer vetenskapliga metoder i ledning av organisationer. Chefer i synnerhet kan då använda en mer vetenskaplig process med formandet av hypoteser och att designa samt utföra experiment för att testa hypoteserna och analysera resultaten innan beslut fattas. En organisation som utnyttjar data på detta sätt fattar beslut baserat på de resultaten av experimenten och fördelarna med den synen på data finns demonstrerat i vetenskaplig forskning. Det är dock inte alltid möjligt att kunna experimentera under kontrollerade former och alternativet är att leta efter naturligt förekommande data att undersöka för att kunna identifiera och förstå variationer. Detta kan sedan hjälpa till att få en ökad förståelse för variationerna inom data och hur effektiviteten kan förbättras. Ett bra exempel kan vara att kolla på variationer av två arbetsplatser som har liknande arbetsuppgifter och titta på om det finns variation i grad av effektivitet och vad som kan vara orsaken till den. Manyika et. al. (Manyika et. al., 2011).

Uppdelning av kunder i segment för att skräddarsy handlingar.

Uppdelning av kunder i segment är ingenting nytt i sig, speciellt när det gäller företag som säljer varor och tjänster direkt till kunder. Dessa företag segmenterar och analyserar kunderna genom olika attribut som exempelvis demografi, inköpsvanor samt beteende och kan med detta göra ta bättre affärsbeslut baserat på omfattande information. Bra exempel på områden som använder sig av kundsegmentering är försäkringsbolag samt kreditkortsföretag då de dagligen sysslar med riskbedömning och därför behöver omfattande information för att kunna göra bra bedömningar och fatta bra beslut. Manyika et. al. (Manyika et. al., 2011).

Ersätta och understödja mänskligt beslutsfattande med automatiserade algoritmer.

Avancerade analyser kan avsevärt förbättra beslutsfattande, minimera risker och underlätta upptäckandet av nya insikter som annars skulle varit dolda. Big Data både tillhandahåller data som behövs för att kunna utveckla algoritmerna och för att algoritmerna skall fungera. Big Data-analyser i dagsläget inkluderar regelbaserade system, statistiska analyser och tekniker för maskininläring så som neurala nätverk. Bra exempel på område där automatiserade algoritmer används flitigt är dagligvaruhandeln som har litet utrymme för lagring och hela tiden behöver kolla försäljningsstatistik och beställa nya varor när de börjar ta slut. Manyika et. al. (Manyika et. al., 2011).

Innovera nya affärsmodeller, produkter och tjänster.

Big Data möjliggör för företag att kunna skapa nya produkter och tjänster, förbättra befintliga produkter och tjänster samt uppfinna helt nya affärsmodeller. Detta har möjliggjort för företag att få bättre feedback från sina produkter, ofta i form av sensorer. Denna feedback kan användas för att exempelvis utveckla serviceerbjudanden för eftermarknaden på produkter eller att använda data som bas för utveckling av nya produkter. Manyika et. al. (Manyika et. al., 2011).

Russom (Russom, 2011) Redovisar en undersökning om nyttan med Big Data där 325 respondenter har angivit 15 olika typer av nytta (se Figur 5 nedan). Följande är de nyttor som angivits av flest respondenter:

Which of the following benefits would ensue if your organization implemented some form of big data analytics? (Select five or fewer.)



Figur 5: Undersökning av uppfattad nytta med Big Data (Russom, 2011 s.11)

Big Data innebär nya utmaningar för organisationer och samhället. Michael & Miller (Michael, 2013) menar att det finns många osäkerheter och utmaningar när det gäller Big Data, men förhoppningsvis överväger de positiva fördelarna och utmaningarna de negativa. Även om Big Data kan innebära extremt stora mängder användbar information, finns det också utmaningar avseende hur mycket data som skall lagras, hur mycket får det kosta, hur skall data säkras? Och hur länge skall den underhållas? (Michael, 2013). Big Data kommer också att skapa nya etiska utmaningar när företag använder Big Data för att lära sig mer om sina anställda, kunder i syfte att förbättra verksamheten. Men dessa förbättringar sker på bekostnad av den personliga integriteten då anställda kontinuerligt mäts på allt de gör. Till det kommer multimedia datamängder som kan göra gränsen mellan det privata och det offentliga otydlig. Ett bra exempel på det är när människor filmar andra människor (avsiktligt eller oavsiktligt) med iPhone, lägger upp det på Facebook, där den personen som blir filmad kanske gör något som denne uppfattar som något mycket privat, och helt plötsligt finns det ute på nätet. McAfee & Brynjolfsson (McAfee, 2012) presenterar fem utmaningar för företagsledningen avseende Big Data:

1. **Ledarskap.** För att lyckas med Big Data menar de att det krävs ett ledarskap som sätter tydliga mål, definierar vad som förväntas och ställer de rätta frågorna.
2. **Hantering av kompetens.** När Big Data blir mer viktig för företaget krävs det att det finns rätt kompetens att få ut nyttan med Big Data. Det gäller också att ha rätta yrkeskategorier anställda. De nämner särskilt Data vetenskapare (eng. Data Scientist) som en viktig yrkeskategori.
3. **Teknologi.** Att det finns verktyg tillgängliga för att hantera de stora volymerna och variationerna samt hastigheten av datagenerering och frekvensen av datahämtning.
4. **Beslutsfattande.** Det gäller att ”placera” informationen och relevanta beslutsbefogenheter på samma ställe.
5. **Företagskultur.** Man skall inte fråga sig: Vad tror vi? utan Vad vet vi? Vilket innebär att man agerar mer på insikter snarare än chansningar.

Det finns sålunda en hel del utmaningar med Big Data som måste hanteras så de inte blir barriärer eller risker för företaget.

Risker och barriärer

Det finns flera risker och problem som organisationer behöver hantera för att kunna få ut den fulla potentialen av Big Data. Manyika et al. (Manyika et. al., 2011) har kommit fram till följande fem problem vilka behöver hanteras:

Datapolicy och personlig integritet

I informationssamhället vi lever i idag digitaliseras samt sprids data över både geografiska och organisatoriska gränser. Detta gör att policies kring data blir allt viktigare och dessa policies inkluderar bland annat sekretess, säkerhet, immaterialrätt, och även ansvarsskyldighet. Det samlas in mer information än någonsin om individer i dagsläget, detta inkluderar bland annat data om individers hälsa samt finansiella situation som individerna kräver att myndigheter och företag skall ha datasekretess kring. Denna typ av personlig data kan ge enorm nytta både för exempelvis val av lämplig behandling på sjukhus och även val av mest fördelaktig avbetalningsplan för ett lån. Personlig data är dock väldigt känslig och därför kan det vara svårt för myndigheter och företag att balansera mellan sekretessen och hur mycket nytta man kan få ut av data. Det ställer ännu högre krav på datasäkerhet för att motverka att data kommer i fel händer. Det finns även

juridiska problem kring data då den skiljer sig väsentligt från andra tillgångar. Data kan kopieras och blandas med annan data utan problem, och samma data kan användas av flera användare samtidigt. Detta gör att immaterialrätten i synnerhet är viktig att tänka på när datapolicys skapas, både ägarskapet och användningen av data men även vad som räknas som laglig användning av data. Vad gäller ansvarsskyldighet är det viktigt att bestämma vem ansvaret ligger på om och när felaktig data leder till negativa konsekvenser Manyika et. al. (Manyika et. al., 2011). Sedan finns det ju en etisk sida av det hela. Boyd & Crawford (Boyd & Crawford, 2012, sid. 671) uttrycker det som att:

”Just because it is accessible does not make it ethical”

De beskriver problemet som kan uppstå när data från sociala media används för analyser mm som inte de aktiva i den aktuella sociala median är medvetna om. Även Bollier (Bollier, 2010) ställer sig frågan vilka etiska ställningstaganden har myndigheter och företag gjort när de använder Big Data riktad mot människor utan deras vetskap. Han menar att här finns stora risker för att den privata integriteten överträds.

Gamla teknologi, metoder, äldre system (eng. legacy systems) samt inkompatibla standarder och format kan hindra dataintegration och avancerade analyser av data vilket gör att man kanske inte får ut den optimala potentialen av Big Data (Manyika et. al., 2011). Detta är även något som Cozzocrea et. al. (Cuzzocrea, et. al., 2011), nämligen heterogenitet hos datakällor och inkongruens från många olika, gamla som nya system.

Organisationell förändring och kompetens är en potentiell barriär för att få ut nyttan med Big Data. Manyika et. al. (Manyika et. al., 2011) menar att om en organisation inte har tillgång till rätt kompetens för att skapa insikter från Big Data, då lär man heller inte få ut nyttan av det. Vidare om man inte har förmågan att förändras i samma takt som konkurrenterna, får man inte heller ut fulla potentialen av Big Data.

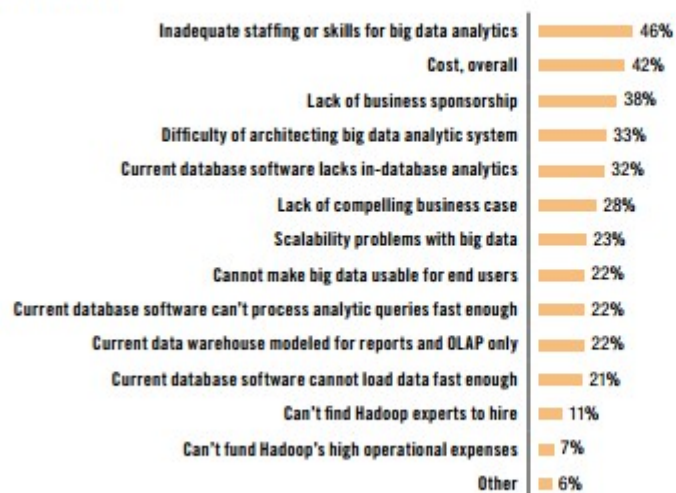
Datatillgång, eller snarare brist till tillgång till viktig data kan vara en stor barriär för att få ut potentialen med Big Data. Det kan vara så att viktiga aktörer inte delar med sig av data, eller att det är stor konkurrens om viss data. Inom en organisation kan det vara så att enheter håller på sin data på grund av att det ger dem en stark maktställning gentemot andra. Men det kan också vara

åt andra hållet. Att stor data tillgång inte alltid leder till ”bättre” data. Boyd & Crawford (Boyd & Crawford, 2012) menar att bara för att forskare får tillgång till stora mängder data innebär inte att metodfrågor avseende analys och sammanställning av data är irrelevant.

Branschstruktur anges av Manyika et. al. (Manyika et. al., 2011). Som en tänkbar barriär. De tar bl.a. upp som exempel branscher som inte är utsatta för konkurrens och där det inte görs jämförelser av prestationer, tenderar att vara långsammare i sin användning av Big Data.

Boyd och Crawford (Boyd & Crawford, 2012) ställer kritiska frågor om Big Data verkligen kommer att hjälpa oss att skapa bättre verktyg, tjänster, produkter etc. Eller om det kastar in oss i nya besvärliga integritetsfrågor som vi inte har tänkt på. De menar också att Big Data innebär ibland att data lyfts upp från sin kontext och då riskerar förlora sin mening. Russom (Russom, 2011) presenterar en undersökning om vilka som är de vanligaste barriärerna till att man inte får ut potentialen i Big data (se Figur 6 nedan). Den största barriären anses otillräcklig kompetens för data analys.

In your organization, what are the top potential barriers to implementing big data analytics? (Select five or fewer.)



Figur 6: Uppfattning om de väsentligaste barriärerna för Big Data (Russom, 2011 s.12)

I det teoretiska ramverket har vi gått igenom begreppet Big Data, dess definitioner och hur det är relaterat till andra snarlika eller närbesläktade begrepp. Vidare har vi kort presenterat andra aspekter och begrepp som ofta förekommer när det gäller Big Data. Denna del utgör uppsatsens konceptuella ramverk. Den andra delen av det teoretiska ramverket beskriver Big Data utifrån dess tillämpningsområden, nytta och effekter, utmaningar och avslutningsvis risker och barriärer.

I nästa kapitel presenteras uppsatsarbetets metodik och tillvägagångssätt.

3. Metodval

Undersökningens syfte har varit att undersöka hur Big Data uppfattas, dess eventuella nytta samt uppfattade risker. I detta kapitel motiverar vi vår vetenskapliga ansats. Våra metodval beskrivs och relateras till alternativa metoder. Därefter beskriver vi hur datainsamling och analys har genomförts med hjälp av intervjuer, tematiska analyser i vår tolkning av data samt en beskrivning av urval för att uppfylla syftet med undersökningen. Slutligen granskar vi våra källor ur ett kritiskt perspektiv med motiveringar till våra val.

3.1. Vetenskaplig ansats

Forskning har olika sätt att se på hur människor, omgivning och fenomen skall studeras. Synsätt på att bedriva forskning definieras som deduktiv, induktiv och abduktiv ansats. En deduktiv ansats utgår från befintliga teorier och prövar dessa teorier genom empiri. En abduktiv ansats kombinerar både ett deduktivt och induktivt förhållningssätt vilket är en kombination av både en deduktiv och induktiv ansats. Induktiv ansats blev vårt val på grund av svårigheten att entydigt definiera ämnesområdet Big Data. Då termen Big Data kan definieras och tolkas på flera olika sätt (se kapitel 1 samt 2.1) så skulle en enskild definition medföra svårigheter att skapa trovärdighet för undersökningen. Definitionen av termen Big Data skulle inte vara aktuell utanför den institution eller organisation som gjorde definitionen. Således var det viktigt för oss att förutsättningslöst undersöka fenomenet. Patel & Davidson (Patel & Davidson, 2011) beskriver en induktiv ansats som ett sätt att följa upptäckandets väg och således formulera egna teorier och slutsatser utifrån den insamlade data.

3.2. Datainsamling

Empirin i vårt arbete har vi samlat in i delar som litteraturstudier och intervjuer. Initialt gjorde vi en litterär studie där vi ville få reda på hur forskningsläget gällande Big Data såg ut inom akademien. För att kunna undersöka ämnet behöver vi veta nuvarande forskningsläge i ämnesområdet Big Data och närliggande ämnesområden som business intelligence samt hur risker hanteras i organisationer. Intervjuerna gjordes som semistrukturerade intervjuer med en intervjumall som utgångspunkt där vi har format intervjufrågorna från vårt syfte och de frågor vi tyckte var relevanta utifrån vår teoretiska kunskap (se kapitel 2).

Vår empiri är en blandning mellan primära och sekundära datakällor. Skillnaden mellan primära och sekundära, skrivna eller verbala, källor beskriver Patel & Davidson (Patel & Davidson, 2011) som närheten till informationslämnaren. Vidare beskriver Patel & Davidson att det är viktigt att veta varifrån källan kommer för att undvika förfalskningar. Våra primärkällor är våra intervjuer där informanten oftast direkt arbetar med Big Data tillämpningar. Resten av våra datakällor är sekundärkällor.

3.2.1. Litteraturstudier

Vi bildade oss en uppfattning om hur forskningsläget ser ut inom ämnet Big Data utifrån perspektivet informatik. För att svara på uppsatsens syfte upptäckte vi att ämnet är starkt integrerat med ämnesområden som business intelligence och hur kunskap genereras inom en organisation. Forskning inom Big Data har byggt mycket på business intelligence, kunskapsgenerering och riskhantering. Vår kunskap hittade vi genom att söka efter litteratur i Universitetsbibliotekets (UB) Summon söktjänst samt Google Scholar. Sökord som vi använde på Summon var "Big Data", "business intelligence" samt "Big Data risks".

Publicerade artiklar i UB har blivit granskade innan de blir sökbara i Summon vilket medför att källan har en relativt hög trovärdighet. Samma princip gäller visserligen för Google Scholar men artiklar i Summon fick företräde då söktjänsterna var snarlika men vi hade tidigare erfarenhet av att arbeta med Summon.

I ett senare skede lades sökordet "DSS" (Decision Support System) till i sökningen med Big Data för att få en bild av forskningen av Big Data inom beslutsstöd. Tanken var att kunna få en uppfattning av nytta som Big Data kan bidra med i samband med beslutsstöd. Efter att ha granskat abstraktet på artiklarna hittades några intressanta artiklar som hade de forskningsinriktningar som eftersöktes. Intressanta nyckeltermen var "research direction" i

kombination med något av nyckelorden ”business intelligence”, ”Big Data” och / eller ”risk management”. Artiklar som uppfyllde dessa kriterier och inte var publicerade tidigare än 2008 används i denna rapport för att beskriva problemområdet samt bakomliggande teori som leder fram till syftet och frågeställningarna. Då Big Data är ett relativt nytt begrepp, finns det en begränsad mängd akademisk forskning publicerad inom ämnet. Den akademiska litteraturen har därför kompletterats med olika publikationer från s.k. expertorganisationer och forskningsinstitut, såsom IDC, McKinsey Global Institute, The Aspen Institute, TDWI Research, för att nämna några. De senare källorna används särskilt för avsnitten om tillämpningsområden, nytta och effekter, utmaningar samt risker och barriärer.

3.2.2. Intervjuer

Intervjun har inte varit konfidentiell och intervjupersonerna har informerats om syftet med intervjun samt blivit tillfrågade om vi kan nämna deras namn i resultatet. Enligt Patel & Davidson (Patel & Davidson, 2011) så kan en intervju ha varierande grad av strukturering och standardisering. Graden av strukturering under intervjun har vi reglerat med hjälp av en intervjumall. Standardiseringen har varit relativt hög i förhållande till vår tänkta kvalitativa analys men vår bedömning är att intervjun fortfarande är relevant. Syftet med mallen har varit att verka som ett stöd till både intervjuaren och intervjupersonen. Den initiala intervjumallen har haft en låg grad av strukturering och standardisering vilket innebär att utrymme har getts till intervjupersonen att svara på frågorna mer berättande med sina egna ord. Vår första intervjumall (v1.0) var ganska bred med 30 öppna frågor. Andra iterationen av intervjumallen (v1.1) var mer detaljerad och specifik med 48 frågor. Utvecklingen av mallens andra iteration motiverar vi med att vår första mall behövde mer detaljerad beskrivning av vissa frågor som kunde relateras mot varandra. Intervjuerna utfördes genom att vi träffade informanterna öga mot öga. Samtalen spelades in och vi transkriberade varje intervju. Transkriberingen gjordes för att vi skulle kunna göra en kvalitativ analys på vårt resultat. Analyserades gjordes med hjälp av en tematisk analys.

3.2.3. Tematisk analys

Initialt funderade vi på att använda oss utav en fenomenografisk metod. Patel & Davidson beskriver (Patel & Davidson, 2011) fenomenografi metoden som att allt insamlat material skrivs ut. Därefter kategoriseras och kodas empirin under t.ex. rubriker. Då syftet med uppsatsen har varit att studera uppfattningar valde vi istället en tematisk analys för att analysera uppfattningar som vi har samlat in genom våra intervjuer. Tematisk metod används t.ex. för att studera uppfattningar inom psykologiska ämnesområden (Braun & Clarke, 2006).

Analysen genomfördes i följande steg:

1. Först läste vi igenom det transkriberade materialet flera gånger för att få en inblick i hur respondenterna har svarat med avseende på frågorna.
2. Därefter kodade vi vår data genom att namnge mönster och teman, som sedan grupperades till 5 olika teman som relaterar till undersökningens frågeställning.

Mönster och teman som vi har valt ut för analysen är enligt nedan lista:

1. Vad är Big Data och hur förhåller det sig till andra liknande begrepp?
2. Vilka tillämpningsområden anses finnas för Big Data?
3. Hur beskrivs nyttan och potentiella effekter av Big Data?
4. Vilka utmaningar finns när det gäller Big Data
5. Vilka risker och barriärer upplevs med Big Data?

Eftersom studien är induktiv, använde vi ingen förutbestämd mall eller schema för analysen. Analys-stegen skedde iterativt vilket innebär att arbetet med att få fram teman inte var en linjär process. Genomläsningen och skrivandet av analysen vävdes samman under iterationernas gång (Braun & Clarke, 2006). Resultatet är vår tolkning av hur informanterna har uttalat sig när det gäller vår frågeställning. Således har vi letat efter mönster av olika kategorier i intervjuernas utsagor och relaterat mönstren mot undersökningens utfallsrum. Det innebär att vi har knutit ihop resultatet av intervjuerna mot den teori som har presenterat i kapitel 2. Slutligen har vi relaterat kategorierna, eller utfallsrummet, till varandra.

3.2.4. Urval av informanter

Holme & Solvang (Holme & Solvang, 1997) menar att syftet med en kvalitativ intervju skall vara att öka informationsvärdet och skapa en grund för djupare och mer fullständiga uppfattningar för det fenomen som vi studerar. Vi vill få ett brett informationsinnehåll för att besvara forskningsfrågorna med en stor variationsbredd. Det innebär att vi vill intervjua personer som är involverade inom Big Data med olika förutsättningar. Således förväntar vi oss att spridningen på definitionen av vårt undersökningsfenomen ökar och intervjupersonerna innehar varierande kunskap om Big Data. Våra informanter har valts ut baserat på att de arbetar inom eller åt en organisation som bedriver verksamhet inom ämnesområdet Big Data. När vi letade efter personer som arbetade inom Big Data kunde vi inte hitta några renodlade tjänstebeteckningar med titeln Big Data. Således har det inte varit ett krav från vår sida att personen uttalat arbetar med Big Data. Det räckte för oss att personen arbetade med ett system eller ämnesområde som ingår i vår beskrivna teori i kapitel 2.

3.2.5. Genomförande

Initialt gjorde vi en litteraturstudie och bildade oss en uppfattning om hur forskningsläget såg ut inom Big Data området. Vi undersökte olika metoder som skulle kunna vara intressanta för att svara på frågorna och vi skapade ett initialt utkast av rapporten. Utkastet agerade även som en guide för vårt fortsatta arbete. Potentiella informanter valdes ut och vi skickade ett e-postmeddelande för att efterfråga en intervju. Totalt skickade vi ut 15 e-post meddelanden varav 10 svarade och vi genomförde 5 intervjuer. Vi använder dock 4 intervjuer i uppsatsen då den sista intervjun genomfördes så pass sent att vi inte kunde analysera den innan uppsatsens inlämningstillfälle. Därefter började vi transkribera intervjuerna och vi tillämpade en tematisk analys för att hitta mönster för vårt resultat.

3.3. Analys av kvalitativ data

Med utgångspunkt från vår empiri, de transkriberade intervjuerna, har vi kunnat hitta mönster i vår data. Vissa teoretiska beskrivningar var återkommande och den tematiska analysen var i stark konkurrens med fenomenografisk metod. För att få en förståelse var vi tvungna att ha ett teoretiskt ramverk att utgå ifrån. Analysen av intervjun redovisas i resultatdelen (kapitel 4) där vi relaterar informanternas utsagor till den teori som har varit återkommande inom Big Data.

3.4. Källkritik

Sambandet mellan datakällor brukar delas i primära och sekundära källor enligt t.ex. Holme & Solvang samt Patel & Davidson (Holme & Solvang, 1997; Patel & Davidson, 2011).

Kategoriseringen av primära och sekundära källor innebär vilken närhet forskaren har till källan. Till våra primärkällor hör våra intervjuer och till sekundärkällorna har vi litteraturen. Litteraturen som vi använder som vår sekundära källa är dels akademisk litteratur men även en del material från företag förekommer. Risker som kan uppstå med källmaterial som kommer från företag och intresseorganisationer är att materialet kan vara partiskt och ”säljande”. Forskningsresultatet kan påverkas genom att en viss produkt eller företag används som standard vilket kan utesluta en objektiv granskning av själva undersökningsfrågan. Den aspekten har vi varit medvetna om när vi har använt litteraturen och har således filtrerat bort rena ”reklamblad” och säljmaterial.

4. Resultatdiskussion

I detta avsnitt redovisar vi våra resultat av undersökningen. Baserat på vårt metodval görs en tematisk analys av våra intervjuer och vi hittar mönster i vår data. Mönstren redovisas som egna stycken och knyter an mot forskningsfrågorna vilka analyseras djupare i efterföljande kapitel.

4.1. Presentation av respondenter

Nedan följer en matris av respondenter som har intervjuats med tillhörande verksamhet och yrkesverksamhet. För att bibehålla anonymiseringen av respondenterna avstår vi från att ingående beskriva arbetsrollen. Respondenternas åsikter är deras egna och de förbehåller sig att eventuella fakta inte nödvändigtvis behöver stämma helt. Det kan även tilläggas att alla intervjuer skedde via personliga möten och företagsnamn har utelämnats för att skydda verksamheternas intressen.

Alias	Intervju utförande	Typ av verksamhet	Yrkesroll
Respondent 1	Personligt möte	Offentlig sektor	Information inom Sjukvården
Respondent 2	Personligt möte	Konsultbolag	Informationsstrateg
Respondent 3	Personligt möte	Konsultbolag	Verksamhetsförändring
Respondent 4	Personligt möte	Konsultbolag	BI / affärs- / tjänsteutveckling

4.2. Definition och syn på Big Data

Respondent 1

Big Data menar respondent 1 består av den tredje och fjärde dimensionen av data (velocitet samt veracitet). Den första dimensionen är till viss del även inkluderad men snabbheten på hur data skapas är viktigast tycker respondent 1. När mycket data skapas så skall det kunna tillföra någon nytta, utifrån det perspektivet kan Big Data definieras. Realtidsanalyserad data av vital information i stora mängder anser respondent 1 är definitionen på Big Data.

Respondent 1 beskriver ett exempel som involverar jättevolymer av data i hans verksamhet. Det finns en sjukdom som drabbar kvinnor och resulterar i att livmodern tas ut. Om man skulle digitalisera dessa 1000 kvinnors livmödrar, som finns i verksamhetens upptagningsområde, så skulle informationen uppgå till 9 gånger Internets informationsinnehåll. Exemplet är Big Data men innehåller ingen realtidsanalys utan handlar främst om stora mängder data. Det menar respondent 1 är två olika synsätt på Big Data. Realtidsanalys och stora mängder data kan båda bidra med nytta men nyttan skall vara omedelbar. Vidare exemplifierar respondent 1 att om en video site på Internet publicerar 72 timmar video varje minut så är det inte Big Data. Däremot om videofilmernas innehåll skall analyseras, då blir det Big Data.

Utöver att data behöver analyseras så måste den även valideras, anser respondent 1. Ett projekt som blev nyttskapande i efterhand är Bild och funktionsregistret som verksamheten arbetar med. Den data som finns i det systemet måste ibland vara validerad. Bild och funktionsregistret har ursprungligen inte varit tänkt som en Big Data lösning men har blivit det då koncept inom Big Data är väl tillämpbara på Bild och funktionsregistret. Analys i samband med stora volymer data och snabba förlopp anser respondent 1 kräva korrekt data. Spridningseffekten av att felaktig data kommer med i analysen kan bli stor. Data som underbygger en analys får inte vara fel när analysen används för att behandla människor. Således är syftet nyttskapande mer viktigt än själva företeelsen Big Data för verksamheten anser respondent 1. Respondent 1 anger ytterligare ett exempel på validitet som en viktig dimension. Om ett labbprov inte skulle bygga på informationskällor med validerad data så skulle många olyckor kunna inträffa. Därav motiverar respondent 1 den viktigaste dimensionen av Big Data, validitet.

Respondent 2

Enkelt uttryckt anser respondent 2 att Big Data är en vansinnig mängd data. Vidare menar respondent 2 även att data i sig vara små data. Det kan vara små dataströmmar, stor mängd data behöver det inte vara. Respondent 2 exemplifierar variationen med att det är inte alla som sitter med datorhallar som Google och Facebook för att administrera data. I vissa fall så kan faktiskt Big Data strömmar vara ganska små och snäva, små rännilar. Big Data, det är ett populärord anser respondent 2.

Respondent 2 nämner att det har funnits stora mängder data tidigare men nu exponeras vi mer för den stora mängden i och med att vi har allt ifrån Internet, till prylar som pratar och lägger ut data strömmar. Sociala strömmar med ostrukturerad data kompletterar tidigare strömmarna med strukturerad data. Respondent 2 använder inte alltid termen Big Data utan menar att det rör sig om ostrukturerad data eller att man kan hitta strukturer i det ostrukturerade. Oftast är det per definition relativt ostrukturerad data jämfört med traditionella informationssystem som har databasscheman och ganska tydliga tabeller med rader och relationer. Respondent 2 exemplifierar att om man tar till exempel logganalyser, som är en del utav Big Data, från stora webb siter eller loggar från sensoriska instrument. Så är klart att loggen i sig har någon typ av struktur, men man får använda filtreringsmekanismer för att göra någonting med detta. Det är filtreringsmekanismerna som respondent 2s verksamhet arbetar med. Man tittar på bland annat dataloggar och olika system som då jobbar mer med öppna datakällor som infogas i NoSQL databaser där man kan ställa frågor mot relativt stora mängder av ostrukturerad information och ändå bygga användbara visualiseringar eller business taggar.

Respondent 3

Respondent 3 menar på att Big Data är ett begrepp som är uppkommet ur flera slags trender och en av dessa trender är explosionen av data på internet med sociala medier samt konsumenters behov av att dela med sig på nätet vilket leder till att företagen vill få grepp om detta. En annan trend är att tekniken har blivit så oerhört billig då hantering av stora datamängder blir billigare och dessutom kan göras med bra prestanda. Företag blir allt duktigare på kompetensutveckling och därmed att jobba med prediktiva analyser samt att förutsäga framtiden. Detta kombinerat leder till Big Data-trenden och även att datalager, beslutsstöd samt verktyg för analys också blir allt viktigare. Själva begreppet handlar inte bara om volymen och det finns flera olika attribut beroende på vem man frågar men den traditionella definitionen är de tre grundläggande Vna:

volume, variety och velocity.

Ett alternativt ord är extrem data vilken är extrem i avseendet att den är väldigt stor eller svårhanterlig då den är ostrukturerad på olika sätt, eller att den strömmar med väldig hastighet vilket gör att aktionstiden blir begränsad. Gränsen mellan det som kallas beslutsstöd (eng. business intelligence, BI), datalager (eng. data warehousing, DW) samt Big Data är flytande.

Kopplat till trenderna bakom Big Data ligger fenomenen internet of things, alla pratande produkter, uppkopplade försörjningskedjor och maskiner. Dessa genererar stora mängder information så därför är gränsen inte knivskarp mellan begreppen data warehouse, business intelligence, data analytics och Big Data. Analytics är ett mer vidgat begrepp av beslutsstöd, Big Data och prediktiv analys, när det slås ihop kallas det business analytics, data analytics eller bara analytics. Open data är ytterligare en datakälla och kan exempelvis vara pratande myndigheter som vill publicera information som är fritt tillgänglig. Det är samma sak där som med tidigare nämnda fenomen, det är stora informationsproducenter och där finns information som kan användas av oftast företag och organisationer för att dra smarta slutsatser. Det underbygger hela bilden av Big Data och den enorma informationsexplosionen som sker.

Respondent 3 använder liknelsen att man nuförtiden kan bryta malm i en malmgruva och få ut tre milligram guld per ton medan man förr i tiden grävde eller vaskade guld och fick fram stora guldklimpar då och då. Metoderna som användes förr i tiden kan liknas med hur företag har gjort med beslutsstöd och BI förr medan det nuförtiden handlar om att ta betydligt större datamängder och använda effektivare verktyg för att få ut värdefull information. Det blir som ett Big Data-raffinaderi. ETL-processer är det som gjorts hittills med beslutsstöd som tankar av operativa system så för att vrida och vända på data. Det är likt guldklimparna och en ganska enkel dataexploatering. Big Data kräver helt annan infrastruktur för att ta enorma mängder data och destillera det ner till någonting som sen kan stoppas in i ett beslutsstöd. Det kanske är ett destillat av x antal exabyte data som har hämtat in vilket sedan bara blir ett Excel-ark med tio rader. Det är en helt annan process än den inom data warehouse men det måste samexistera då den destillerade produkten måste sättas i ett sammanhang som finns i ett beslutsstöd. Respondent 3s tankar kring data relaterat till internet of things är att företag kan ha uppkopplade försörjningskedjor dvs. att företagen kopplar upp hela sin distributionskanal, alla affärspartners, underleverantörer och deras produktionskedjor så att alltihopa kommunicerar. Detta gör att man snabbt kan få en uppfattning av problem hos bl.a. underleverantörer och hur det kommer påverka hela produktionskedjan samt hur man kan jobba prediktivt för att motverka problem. Respondent 3 tror att detta är en viktig del av Big Data för att kunna smörja och effektivisera hela produktionskedjan.

Respondent 4

Respondent 4 väljer att definiera Big Data enligt de klassiska V'na med stora, snabba och föränderliga data som man inte riktigt kan hantera på samma sätt som man gjort med data tidigare. Han menar på att vi är vana vid att om vi har information så sätter vi upp en datamodell och stoppar in data i en relationsdatabas vilket han menar räcker ganska långt idag. Det kommer mer och mer sensordata samt webbtrafik och när det är så enorma mängder data så funkar det inte att stoppa ner i en relationsdatabas. Det behövs således andra tekniker för att hantera data. Lägg därtill att det bara är den tekniska sidan, med alla tillämpningar som är relevanta inom businessvärlden så börjar man fundera kring vad man kan göra med all information och respondent 4 menar på att det kan förändra världen på olika sätt. Det kommer inte behöva gissas lika mycket utan man kommer med data kunna ta reda på hur det mer detaljerat ligger till när man undersöker saker. Speciellt då det ofta handlar om att fråga vad folk tycker och att detta ibland går isär med vad folk egentligen gör. Respondent 4 jämför med beteendevetenskaplig forskning där vissa analyserar väldigt mycket teoretiskt och vissa andra tittar på experiment och vad som händer, och där det då visar sig att folk väldigt ofta säger en sak och gör en annan. Respondent 4 väljer även att definiera Big Data som att det är större än att bara hantera de stora datamängderna och kunna analysera data vilket många från respondent 4s bransch definierar det som.

Data warehouse är ett traditionellt sätt att lagra och strukturera data för analytiska ändamål. Utifrån respondent 4 definition av Big Data kan man säga att det inte är så ofta man trycker ner Big Data i datalager. Han menar på att Big Data ligger lite mer parallellt med data warehouse. I ett data warehouse så har man strukturerad data som man har koll på och vad gäller Big Data är det ett helt annat fokus. När man pratar om finansiell rapportering skall allt stämma på kronan medan det kring Big Data är lite annorlunda. Där finns fel och kvalitén på data är lägre för man har så mycket information att man dränker felen. Det hade aldrig en redovisningsekonom accepterat för där handlar det mer att det skall stämma på kronan på kontot. BI för respondent 4 är verksamhetsstyrning, att ha koll på sin verksamhet och använda informationen som finns i verksamheten. I BI är data warehouse en del och på det tillkommer rapporter och analyser. Respondent 4 menar på att Big Data inte kommer ersätta BI utan att det är lite mer likt två sidor av samma mynt där ändamålen skiljer sig.

Respondent 4 syn på open data är där finns möjligheter att använda informationen. Speciellt då utomstående personer kan sitta och analyserar data och kanske hitta något intressant. Problemet som respondent 4 har upplevt i sina projekt är dock att man vill ha informationen på ett ganska specifikt sätt och när man hittar öppen data så får man den ofta på en lite högre struktureringsnivå

och man kan då ha tappat möjligheterna att använda den i sammanhanget. Den öppna data hade varit lättare att använda om den hade funnits på en lägre nivå av strukturering så den kan bearbetas och användas för ens egna ändamål.

Internet of things menar respondent 4 kommer leda till enorma mängder ostrukturerad data vilket gör att Big Data och internet of things är väldigt tätt sammankopplat. Det börjar börjar sättas ut någonting kallat i-beacons i butiker så företagen direkt, när kunden går med sin telefon kan ta emot erbjudanden och liknande i realtid samt kolla vart kunderna befinner sig i butiken. Detta gör att det genereras ofantligt mycket mer information som respondent 4 menar växer nästan exponentiellt. Ofta om man tittar på data warehouse och BI samt databaser så växer den kapaciteten linjärt och därför måste det till nya sätt att hantera data. Quantified self tycker respondent 4 är väldigt spännande och nämner att inom några år kanske man kommer kunna läsa av kroppsvärden i sin klocka samt att man skulle kunna använda all den informationen till att ta fram exempelvis läkemedel. Respondent 4 hade läst någonstans att en stor del av alla sjukdomar människan lider av hade vi kunnat slippa bara genom att äta och motionera regelbundet. Då skulle man kunna tänka sig att om man kopplar upp Big Data och quantified self hade man kunnat ha en form av personlig coach hela tiden vilket hade tagit över mycket av läkemedelsbranschens roll men att läkemedelsbranschen då skulle kunna lägga mer energi på att hitta botemedlet för cancer, malaria eller liknande. Respondent 4 nämner att det i branschen har pratats om att Big Data skulle kunna lösa alla problem i världen men att man kommer inse med tiden att det inte löser allt men kanske kan underlätta vissa problem. Precis som internet of things så handlar social media också om enorma datakvantiteter och genom att koppla ihop de tre fenomenen: internet of things, quantified self och social media så kan man få ut väldigt mycket intressant information.

4.3. Nyttan och effekter

Respondent 1

Enligt respondent 1 så har dennes verksamhet 1500+ IS/IT system i vården som ibland är väldigt autonoma. Om de skulle kunna sammanföra den informationen, inte bara till ett system utan sammanföra informationen så de skulle kunna skörda informationen kring en patient ur patientgruppen. Då skulle verksamheten kunna använda det till ett beslutsstödssystem.

Respondent 1 menar att verksamheten inte är där ännu. De flesta beslutsstödssystem som verksamheten har är inte kopplade till den information som finns i just det systemet. Med det menar respondent 1 att om man har ett labbsystem som upparbetat information så är beslutsstödet byggt utifrån den källans information. Men beslutet kan ha sett annorlunda ut om man hade haft tillgång till ett annat system. Därför ligger mycket utav beslutsstödet idag hos individen. Hälso och sjukvårdspersonalen som gör en aggregering av, det här ser respondent 1 i IS/IT systemet, data och information vilken kommuniceras ut till patienten. Beslutet menar respondent 1 fattas baserat på individens informationskälla. Det vill säga den beprövade erfarenhet, bakgrund och tidigare kunskap. Vidare exemplifierar respondent 1 Bild och funktionsregistret som har börjat användas väldigt mycket med tillhörande forskning kring det systemet. Respondent 1 anser att dennes verksamhet har tagit upp Europas största forskningsprojekt inom vården där en utav IT faktorerna var att infrastrukturen, som klarar att hantera Big Data, redan fanns på plats. Forskningsprojektet heter SCAPIS, en förkortning för Swedish Cardio vAscular Pulmonary Image analysiS där man screenar folk. Ett sådant projekt anser respondent 1 ha potential men är samtidigt ganska skrämmande ur ett individuellt perspektiv.

Screeningen innebär att man förebygger, patienterna behöver inte vara sjuka utan de väljs ut slumpmässigt och respondent 1 tror att det är 30000 individer som väljs ut vilka skall följas under 25 år. I projektet är det väldigt mycket utav labb, kardiologi, radiologi, egen utsaga hos försökspersonerna. Med hjälp av de personer som vill ställa upp på den här vetenskapliga studien kommer verksamheten definitivt att använda sig utav de begrepp som förekommer inom Big Data. Projektet kommer kanske ligga till grund för framtidens beslutsstödssystem menar respondent 1.

Respondent 1 gör en jämförelse mellan den egna verksamheten och den industriella verksamheten eller företagssidan. Där tror respondent 1 att den industriella sidan är längre gånga idag i sina Big Data satsningar på grund av ett lite mer svart eller vitt synsätt. Ett exempel som respondent 1 tar upp att ett snabbt förlopp som kan hända i den egna verksamhetens informationskälla är att någon twittrar om någonting. I den egna verksamhetens värld skulle en

patient kunna Twittra om sig själv, men det skulle inte vara så genomslagskraftigt och skapa badwill på något stort sätt för verksamheten tror respondent 1. Men om en leverantör har släppt ut en dålig eller bra produkt och det Twittras om den så påverkar ju det leverantören på något sätt. Det kan ju vara att det är positivt då skall leverantören ha marketing som fångar upp det och kör den virala succén. Är det något negativt så skall leverantören vara proaktiv och bemöta kritiken. Ett viralt förlopp på sociala medier med stora mängder data sker ögonblickligen. Förloppet är ändå bara en liten ström utav allting annat som händer på Internet. Respondent 1 tror att än så länge så är den streamande datamängden, än så länge mest konsumentdrivet. Respondent 1s verksamhet är däremot öppna för möjligheten att i framtiden kommer patienterna kunna ta mer ansvar och få kunskap att aktivt delta i de sociala dataströmma. Att patienterna får möjligheter, kommer att bli motsvarande sociala medier idag.

Sen har vi SCAPIS det är utan det här konceptet (Big Data) så hade verksamheten inte kunnat köra det projektet på det sättet anser respondent 1. Verksamheten har ett annat projekt som är tidig upptäckt utav cancer. Projektet baseras helt på koncept inom Big Data därför att nu kan verksamheten i realtid analysera det som händer i verklighet. Respondent 1 menar att analysen skiljer sig gentemot tidigare sätt att analysera data då har man tidigare har fått göra analysen rekonstruktivt. Det innebär att man har fått vänta tills blodproppen har ägt rum, kartlägga var det finns någon källa av information där dom här kriterierna för proppen har mötts och sedan insamla data och där göra analysen. Med Big Data konceptet kan verksamheten istället sätta kriterierna före händelsen har inträffat och sedan identifiera händelserna i realtid. Verksamheten tillämpar Big Data koncept men ser inte vad resultaten ger än.

En av de första tillämpningarna som kunde möjliggöra tidiga upptäckter av anomaliteter innehöll 30 000 individer, 24 000 ur kontroll och hittade 6 000 positiva fyndigheter. Respondent 1 beskriver en av personerna blev rörd till tårar när forskningsledaren gick härifrån för att hon hade sett att det var ett massivt arbete att ta fram det här underlaget. Resultatet fick forskningsledaren med sig det på en minnessticka och kunna ägna sig åt kärnforskningen direkt. Ett liknande fall på nytta identifierar respondent 1 är vid diskussioner med läkemedelsföretag. Det har en stor genomslagskraft för verksamheten om de har avsatt 100 miljoner för forskning, klinisk forskning, så har de räknat med att ca hälften går åt till upparbetning av det material man skall forska på. Det här menar respondent 1 påverkas av tillämpningar av Big Data när möjligheten uppstår att budgeten kan fördelas som att 30 miljoner som går åt till upparbetning och 70 till kärnforskningen. Det menar respondent 1 gör att man har höjt det till en högre nivå med hjälp av Big Data. En annan sidoeffekt, som respondent 1 tror på, blir att fler aktörer intresserade av att

hjälpa till att göra satsningar inom forskningsområdet.

Slutligen menar respondent 1 att vi är för tidigt ute för att specifikt peka på ett resultat utav deras Big Data satsning. Men han är väldigt komfortabel med att det kommer beroende på det konceptuella kring Big Data och vad det har givit oss.

Respondent 2

Syftet för att fånga en dataström anser respondent 2 är en viktig faktor. Saknas ett syfte så spelar det ingen roll om det finns en logg på en petabyte om vi inte vet vad vi skall ha den till. I det fallet menar respondent 2 att det kanske räcker med ett Excel-ark som är semikolonseparerat. Respondent 2 menar att Internetföretag ofta justerar sin verksamhet baserat på analyser i realtid. Nyttan är att om organisationen har sensoriska möjligheter att ta in data så bör den användas för att forma organisationens verksamhet. Ju bättre möjligheter som organisationen har på att ta in data desto bättre kan organisationen kalibrera sina tjänster gentemot sina kunder. Vidare menar respondent 2 att kunskap som kan identifieras med Big Data är att hitta anomalier i beteendemönster. Detta kan göras till exempel i form av loggar. Sådant som inte blotta ögat kan se. Det handlar om att tvätta data och kunna visualisera den på ett enkelt sätt. Visualiseringen kan sedan användas för att hitta konstiga beteendemönster på sin site och utveckla den baserat på anomalier.

Ett annat tillämpningsområde är att kunna matcha olika dataset och hitta nya tillämpningar. Datakällans ägare kanske inte kan styra hur sin data skall användas då andra kommer att nyttja denna funktion. Exempelvis så kan tidtabeller för båtar som går under öppningsbara broar användas för att räkna ut vilken väg som är bäst att ta sig över ett vattendrag.

Respondent 3

Respondent 3 nämner att dennes verksamhets Big Data-satsning startade med ett gäng konsulter som tyckte att Hadoop var spännande så dom satte upp ett kluster för att se vad det fanns för möjligheter med Big Data. Initiativet kommer oftast på kund och marknadssidan där organisationen har en vag uppfattning om att dom vill veta mer om vad som händer bland kunderna. Detta området brukar benämnas customer insights eller customer analytics. Bakom detta finns det oftast ett rejält behov av att få grepp på sociala medier och med detta börjar kolla på Big Data på en mer it-strategisk nivå. Verksamheten jobbar mycket med Microsofts produkter

för att använda MPP (Massive Parallell Processing) vilket är en förutsättning för att kunna hantera stora transaktionsvolymer.

Vad gäller hur en Big Data lösning kan se ut så finns ett exempel i dagsläget med en kund som är i ett tidigt stadium av implementeringen där arkitekturen håller på att byggas upp. I det projektet så finns det en Hadoop arkitekt, en datalager arkitekt och en beslutsstöd/utvecklar arkitekt.

Respondent 3 påpekar även att det är viktigt att kunden har kompetensen för att kunna ta tillvara på resultatet. Det är långt ifrån bara den tekniska biten som behövs och kompetensbehovet av att kunna analysera och använda resultatet skall ej underskattas.

När det gäller områden som kan ha nytta av Big Data nämner respondent 3 fordonsbranschen samt även e-handel och detaljhandel där prissättning kommer bli en allt vanligare konkurrensfaktor. Hela detaljhandelssektorn har ett starkt konsumentled där man använder konsumenterna för att gratis eller mot viss ersättning tillhandahålla information till företagen på olika sätt. Konsumenterna är villiga att dela med sig av allt möjligt för en struntsumma. Respondent 3 tror även att myndigheter kommer att titta närmare på Big Data, även kommuner kanske vill få ett grepp om vad kommuninvånarna tycker och vad som sägs bland medborgarna. Organisationer som är i nära kontakt med kunder kommer dra nytta av Big Data i framtiden enligt respondent 3.

Framtiden kan vara skrämmande då bl.a. Netflix har baserat tv-serier på vad konsumenterna vill se och respondent 3 tror att den värsta möjliga bilden av framtiden är att vi hänger i form av någon Matrix (referens till filmtrilogin Matrix) där vi blir matade med den informationen vi vill och aldrig blir utmanade, en tråkig värld som är tillrättalagd efter den stora massans behov och önskningar. Även om framtiden är oviss, märker respondent 3 ett tydligt behov av Big Data satsningar ute hos kunder och spår att Big Data trenden kommer bli stor när den väl slår igenom.

Respondent 3 nämner innovation som en viktig komponent och menar på att det med hjälp av Big Data kan skapas nya affärsmodeller och produkter vilket driver marknaden framåt. Han nämner Spotify som ett exempel på en ny produkt som uppkom i Internet eran efter CD skivornas tid. Den största fördelen med Big Data enligt respondent 3 är ökad kunskap om kunden samt deras önskningar, och då möjligheten att som företag kunna anpassa sig efter behovsbilden. Det kan vara svårt att uppskatta vart värdet finns i en tidig fas och om man ska återkomma till liknelsen med gruvdriften så förekommer det att man deponerar gruvmassor för att man tror att man i framtiden kan ha teknik som gör att man kan utvinna ännu mer ur massorna, det kan vara liknande när det gäller data också. Man kanske inte ska kasta data då man inte ser värdet av den idag. Om några år kanske den är ovärderlig.

Respondent 4

Respondent 4 nämner att dennes verksamhet ej har en tydlig Big Data-satsning för tillfället utan att dom istället satsar mer på områden och att ett fokusområde är marketing eller customer analytics och marketing automation. Där handlar det i vissa fall om företag med tiotusen kunder och då blir det inte lika mycket data, till skillnad från exempelvis banker som har mycket mer kunder och där det då blir väldigt mycket data. Respondent 4 förklarar att konsultbolaget han jobbar på är väldigt kunddrivna och hela tiden sitter och känner av vart vindarna blåser för att kunna vara med och hoppa på trender när de väl uppkommer. Storskaliga satsningar på Big Data är det dock inte många som håller på med i Sverige ännu. Respondent 4 tror att det är på gång men att det är fortfarande ganska nytt.

Respondent 4 har en bild av att Big Data fortfarande är ganska mycket fokus på det tekniska. Ska man göra Big Data lösningar behöver man satsa mycket på tekniken och det är fortfarande en ganska svårtillgänglig teknik. Man kan ta ett exempel med Hadoop. Ska man sätta upp en Hadoop lösning behöver man folk vilka är väldigt skarpa på programmeringsspråket Java och man behöver ha en bra koll på teknikerna och systemen kring Hadoop. Det var liknande inom data warehouse och vanliga databaser för tjugo år sen men idag kan nästan vem som helst klicka next, next, next och installera så har man en fungerande databas som man kan börja använda. Där är en ganska stor teknisk barriär fortfarande. Säg att man ska köra ett Big Data-projekt, mycket av kostnaden tror respondent 4 skulle gå åt till att sätta upp systemen om man tänker en stor lösning, då behöver man investera ganska mycket i infrastruktur. Respondent 4 resonerar även kring kompetens ur ett analysmässigt perspektiv där man brukar prata om data scientists, folk som kan hantera information och statistik. Respondent 4 menar att på ett sätt kommer det bli viktigare men på ett annat sätt kanske det inte behöver bli viktigare då det kommer till ett läge där utvecklingen har gjort lösningarna enklare så att vem som helst skall kunna hantera dom. När det gäller väldigt avancerade analyser på data krävs det som tidigare nämnt, folk som har bra koll på statistik, speciellt om det är analyser som skall användas till något kritiskt istället för experimentellt, exempelvis inom sjukvården. Tänker man ett steg längre kan det ibland vara mer fördelaktigt att få en fingervisning istället för komplett statistik. Det är dels är en kostnadsfråga att komplett statistik kräver mer omfattande analys och därmed kostar mer pengar, men även att man inte alltid behöver den säkerheten som en statistikdoktor skulle erbjuda utan det räcker ganska långt med tumregler. Respondent 4 exemplifierar med att om man ser att när priset sänks på varmkorv så får kunden fler besökare i butiken och då säljer kunden exempelvis fler paraplyer. Denna fingervisning räcker ofta i sammanhanget och man behöver inte veta med en halv procents

konfidensgrad att det verkligen är så.

Respondent 4 tror att dataintensiva branscher har mycket att hämta i och lätt att komma igång med Big Data. Även företag där processer många och mycket är IT baserade så man har någonting att utgå ifrån. Datadriven affärsutveckling har förändrat ganska många branscher under en längre tid. Mer vad gäller nytta och effekter kan vara kunskap och insikter där det egentligen är samma sak för Big Data som med klassiska beslutsstöd och analys, att man utmanar gamla sanningar. Respondent 4 tycker det är väldigt roligt att testa hypoteser som man har om verksamheter, både roligt när dom stämmer och någonting är på ett sätt som alla tror att det inte är, men även för att validera.

4.4. Risker, barriärer och utmaningar

Respondent 1

Respondent 1 menar att det finns främst två risker med Big Data. Den första är informationsövertag eller upplevelsen av ett informationsövertag samt rädslan att någon annan skulle få det. Den andra risken är att inte veta vad Big Data är och vad det skall användas till.

En av utmaningarna menar respondent 1 är att någonstans skall man förklara för sina patienter varför verksamheten samlar in information om dem och vem som får ta del av informationen. I slutändan är det patienten som bestämmer. Oftast är det inte något problem om datainsamlingen presenteras det på ett korrekt sätt. Fenomenet menar respondent 1 kommer att leva under en längre tidsperiod. Men å andra sidan ju fler som är med, det är desto lättare försvinner individen i datamängden. Beroende på hur verksamheten skördar nyttan så kan man faktiskt komma fram till att patienten har en sjukdom som är väldigt ovanlig. Då tittar man utifrån patientens sjukdomsbild istället när anomalien väl är upptäckt. Risken finns då att man får ett annat svar då patienten framträder väldigt tydligt som individ och det är kanske något man, som patient, inte har tagit ställning till när man sade ja till att data sparas och analyseras.

Det blir ju lite utav en organisatorisk risk påverkar organisationens trovärdighet menar respondent 1. För om det börjar sprida sig bland patienter, att organisationen använder data till annat än det som var tänkt, så sjunker ju förtroendet för organisationen. Då kan organisationen stå inför scenariot att flertalet patienter inte vill delta med sin data, menar respondent 1. Trust, eller pålitlighet, i data är en viktig faktor för att få patientens förtroende menar respondent 1. Då blir validiteten därefter och ett litet moment 22 uppstår när det gäller organisatoriska risker.

Respondent 1 menar att det kan vara en risk på flera olika sätt om man inte har kunskap om vad Big Data är men tror sig ha kunskap och sprider den felaktiga kunskapen. Visst går det att använda en information på ett visst sätt där det inte är heltäckande. Därför tror respondent 1 att man skall ha med sig någonstans att en risk som är generell är en användare utav ett Big Data kluster blir mindre källkritisk. För ju mer man hittar av det som eftersöks och som tilltalas desto mindre källkritisk blir man, menar respondent 1.

Respondent 1 exemplifierar att det kan ju vara som så att man förlitar sig på ett system, så kan ett litet fel göra att man har blivit så loj att man inte upptäcker felet. Mobiltelefonen är ett exempel som gör respondent 1 paralyserad när den inte fungerar. En dag fungerade inte servern som har hand om e-post och kalendersystemet. Respondent 1 visste inte vart han skulle vara. Förr i tiden kunde han ha mycket i huvudet och visste ungefär vilka möten som fanns bokade. Nu har det varit så att man har vant sig vid att det alltid skall fungera. Ett stort avbrott kan då upptäckas, men om den hade gjort något litet. Säger att var tionde bokning inte finns, ja då hade man troligen blivit varse efteråt någon gång menar respondent 1. Ett sådant här läge kan inte finnas i verksamheten menar respondent 1. Det finns inget efteråt för det kan sluta med att en patient dör när en informationskanal inte fungerar vid ett visst tillfälle. Respondent 1 menar att vi är vana att allting bara fungerar. Ju mer tillförsikt vi får till systemet desto mindre uppmärksamma blir vi menar respondent 1. Så därför tror respondent 1 på kunskap inom tillämpning av Big Data. Först definiera vad Big Data är, hur det skall användas, mätbar nytta samt identifiera de risker som finns. Slutligen menar respondent 1 att man alltid skall fortsätta vara källkritisk.

Respondent 2

Respondent 2 beskriver att lyssningsinstrument, som samlar in stora data, står ofta utanför organisationen och lyssnar beroende på vilken agenda organisationen har. Data samlas ofta in för att göra analyser och få information vilket i sin tur används för att kunna forma sina tjänster och produkter beroende på hur de uppfattas. Utmaningen ligger i att kunna knyta an det ostrukturerade dataflödet till något användbart med verktyg som kan göra textanalyser. Quantified self har vissa lager som man vill hantera själv som t.ex. förbättrad löptid eller vilka sjukdomar man har, säger respondent 2. Vissa data vill man ha öppet och annat vill man inte skall komma ut då det kanske är personligt. Olika data har således olika syften vilket medför att man kanske inte vill visa allt, menar respondent 2. Organisationsmässigt tycker respondent 2 att man

inte kan vara chef för Big Data, utan man måste veta tillämpningen av Big Data satsningen som gör nytta för verksamheten. Man måste förstå de mönster som man skall lyssna på menar respondent 2. Det finns ett organisatoriskt perspektiv som måste förstås och knyta an verksamheten mot de dataströmmar som finns tillgängliga. Respondent 2 anser att man tror att utmaningar inom Big Data är en teknikfråga, men det är det inte. Det är snarare så att om det finns en möjlighet att följa en dataström, skall man ställa sig frågan om det är politiskt korrekt att göra det. Respondent 2 exemplifierar med NSA som lyssnar på all trafik som godtyckligt anses vara av intresse. Ett annat exempel är att om feedback från den som samlar in data, motiveras den avlyssnade till att bli avlyssnad, så blir det mer troligt att gå med på att bli avlyssnad tror respondent 2. En annan utmaning är att kunna ställa rätt och smarta frågor och få ut en korrekt analys av data tycker respondent 2. Respondent 2 ställer frågan att om man börjar övervaka diabetespatienter, kommer vi att ändra organisationen baserat på vad vi får in för data? Om nej då kan vi strunta i att över huvud taget samla in data anser respondent 2.

Respondent 3

Respondent 3 tycker inte att den tekniska biten är något problem egentligen utan berättar att det är kostnader som sätter gränser för hur stora satsningar företag kan göra kring Big Data. Däremot är det svårare när det kommer till sanningshalt och värdet av informationen som fås fram. Värdet kan ej bedömas på samma sätt med någon teknik utan respondent 3 menar att den mänskliga inblandningen för att bedöma värde är viktig. Det finns även en utmaning i att få in Big Data i den strukturerade världen med beslutsstöd men samtidigt kunna låta den utvecklas i en fri miljö vilken främjar innovation.

Respondent 3 menar även på att Big Data fortfarande är ett outforskat område där de som börjat satsa räknas till early adopters och då lärdomarna kring området är begränsade finns det risk att fel slutsatser dras av den insamlade data vilket kan leda till felbeslut som kan skada företaget pengamässigt. Det finns även en risk att man inte kommer fram till nya innovationer med den insamlade data men där betonar respondent 3 att det kan ligga en ännu större risk i att inte göra en Big Data satsning då det kan finnas potentiellt ovärderliga innovationer man då går miste om.

Respondent 4

Respondent 4 resonerar vidare kring organisatoriska barriärer och att upptäcka att saker inte är på ett sätt som man trodde. Oftast är resultatet av analyserna att det leder till någon typ av förändring och det kan vara svårt för kunden man jobbar mot att hantera. Man har i alla år jobbat på ett sätt och sen kommer det plötsligt en analytiker och säger att man borde jobba på ett annat sätt men det vill kunden kanske inte göra för kunden har alltid gjort på sitt gamla beprövade tillvägagångssätt och det har alltid funkade bra. Det handlar mycket om tillit och kan vara en typ av barriär.

Respondent 4 menar på att metoderna för att kunna hantera Big Data finns men att utmaningarna mer ligger i att få ut värdefulla insikter som kan användas. Han nämner även integritetsaspekten av data och tror att det i framtiden kanske leder till att man accepterar att företag vet om mycket vad man gör och när man verkligen inte vill att någon skall ha koll så gör man inte sakerna uppkopplat på nätet. Risker finns även i automatiseringen av verksamheter fast respondent 4 tror att det behövs hittas en balans mellan det automatiserade och människor då människor har en annan form av analytisk förmåga och kan komma till insikter och värdera logiska samband vilka automatiserade system inte kan i dagsläget. Det kan vara en fallgrop och nackdel att man tittar för mycket på data och hitta ologiska samband. Gör man det tillräckligt länge kommer du hitta samband som inte finns. Respondent 4 exemplifierar detta med att några analytiker i USA hittade en korrelation mellan hur mycket ost som såldes och hur många som dog av att trassla in sig i lakanen på natten under en period på tjugo år. Kurvorna följde varandra men det verkar ganska osannolikt att dom har ett samband och med det menar respondent 4 att det är viktigt att veta vad syftet med analysen är för om man bara analyserar och hittar massa falska insikter som man fattar beslut på kan det bli väldigt fel.

5. Analys

Vi resonerar kring hur respondenter har svarat under intervjuerna samt vilka faktorer som har spelat roll i deras svar. Svaren och resonemangen bryts ut och generaliseras. Även våra valda metoder diskuteras och resoneras kring för att belysa alternativa förhållningssätt till vår data. Diskussionen om metodvalet är en djupare analys till metodvalet som har belysts i tidigare metodstycke kapitel 3.

5.1. Diskussion

I vårt teorikapitel (Kap 2.1) har vi redovisat flera sätt på hur Big Data kan uppfattas. De är dock inte svenska eller verksamma inom Sverige i någon större utsträckning. Flera personer verksamma med Big Data (Manyika et. al., 2011; Dumbill, 2012; Russom, 2011) beskriver Big Data som en mängd data vilka är för stora att hantera för vanliga processer eller verktyg som t.ex. databaser. För att hantera problemet föreslås flera lösningar där man kategoriserar och strukturerar data med hjälp av dimensioner, de tre Vna och processer som textanalys. NoSQL databaser som kan göra textanalyser och skapa datastrukturer on-the-fly är några av de praktiska verktyg som kan användas med processerna.

5.1.1. Vad är Big Data och hur förhåller det sig till andra liknande begrepp?

Under intervjuerna har samtliga respondenter nämnt att Big Data fortfarande är ett relativt nytt område vilket inte riktigt har slagit igenom ännu. Förfrågningar från kunder och organisationer börjar dock komma i allt större utsträckning och framtidsutsikten för Big Data verkar ha en potential. Mycket av själva definitionen av Big Data är sprunget av ett behov i en verksamhet som är på gränsen till att nyttja potentialen hos Big Data.

Stora mängder data ha lagrats tidigare i databaser och data warehouses (DW) (Turban, et. al., 2011). Där data warehousing hade rollen som en sorts lagring och möjlighet att skapa data inom en organisation en längre tid så har Big Data ett annat perspektiv. Till exempel data warehouse lösningar lagrar tvättad data som har gått igenom ETL processer, vilket även styrks av Turban et. al. (Turban, et. al., 2011). Den främsta skillnaden mellan Data Warehousing och Big Data måste ändå vara att Big Data använder semi- och ostrukturerad data till skillnad från den strukturerade data som finns i ett data warehouse.

5.1.2. Vilka tillämpningsområden anses finnas för Big Data?

Djicks (Dijcks, 2012) skriver att en organisation kan skapa ny kunskap om sin verksamhet genom att analysera fler olika datakällor. En av respondenterna nämner att integration av hela produktkedjor, vilket knyter an till den teoretiska biten om att skapa datatransparens. Med ökad åtkomst till information inom hela organisationen kan det då fattas bättre beslut grundat på mer information. Mer datagrundande beslut kan ge nytta både hos beslutsfattare som tar strategiska beslut om organisationen och även nere på den operativa nivån där informationsutbyte mellan delar av en organisation kan göra produktionskedjan effektivare genom snabbare informationstillgång.

5.1.3. Hur beskrivs nyttan och potentiella effekter av Big Data?

Värdet av data kan vara svårt att bedöma med teknik och att det krävs mänsklig inblandning för att kunna bedöma värde och samband i data. Ett annat perspektiv utvecklar detta vidare genom att betona vikten av att kunna hitta relevanta samband i data och nämner att man behöver kolla på om sambanden är logiska eller om de bara beror på slumpen eller andra faktorer. Vidare nämns även vikten av att hitta logiska samband och berättar en historia om hur en tidningsskribent vilken skrev regeringskritiska texter, blev besökt av myndigheterna då de hade funnit samband mellan skribenten och sökningar på inhandlande av en tryckkokare som möjligtvis kunde tillverka bomber då det även hade sökts på bombdåd. Detta visade sig helt ologiskt och taget ur luften då det var frun i huset som sökt på och funderat på att köpa en riskokare. Det visade sig även att det var sonen som av nyfikenhet hade sökt på bombdåd kring en aktuell händelse. Med all denna data från samma IP adress i ett hushåll drog myndigheten slutsatsen att det möjligen pågick bombtillverkning då det i själva verket var helt orelaterade samband.

5.1.4. Vilka utmaningar finns när det gäller Big Data

Big Data är fortfarande ett nytt fenomen som fick ett uppsving först 2012 (Kap. 1.1). Det medför att fenomenet inte är fullt definierat och kartlagt ännu. Men det är å andra sidan inte ämnesområdet business intelligence enligt Turban et. al. (Turban, et. al., 2011). En av utmaningarna är således hur Big Data skall kunna visas tillföra ett värde. Svårigheterna med att definiera och avgränsa Big Data (Bhatia, 2013; Pflugfelder, 2013; Chen, et. al., 2014) medför att varje individ eller organisation som arbetar med det skapar sin egen definition. Både Ohlhorst (Ohlhorst, 2013) och Turban (Turban, et. al., 2011) menar att det kan vara problematiskt ur ett tekniskt perspektiv då samma ”Big Data”, eller business intelligence system, inte kan användas mellan olika organisationer utan måste anpassas. Utmaningen bekräftas i denna studie som visar Big Data satsningar som nyttjar konceptuella ramverk (t.ex. Big Data dimensioner, se Kap 2.2) snarare än tekniska lösningar som är sprungna från ett verksamhetsbehov.

5.1.5. Vilka risker och barriärer upplevs med Big Data?

Svårigheterna med datapolices och ägarskap av data diskuteras av flera verksamheter. Anledningen är att datamängden går över både geografiska och organisatoriska gränser vilket gör det svårt att bedöma vilken data som ägs av vem. Detta kompliceras ytterligare då organisationer kan ge någon form av incitament mot att individer delar med sig av data samt även med det faktum att data kan kopieras och spridas till den grad att det kan vara svårt att spåra ursprunget. Återkommande åsikter är att datakällan måste vara pålitlig och validerad. Frågan är då vart ägarskapet av data ligger när individen har fått en form av incitament då det inte längre handlar om att frivilligt tillhandahålla data utan mer blir likt ett utbyte av varor, exempelvis att erhålla rabatt mot kunddata vilket är vanligt förekommande i många butiker idag.

6. Slutsats

Som vi berättade i föregående kapitel så är Big Data inte en helt ny företeelse. Utan det är först nu man har börjat kunna skapa nytta med hjälp av Big Data och dess koncept. Vår forskningsfråga är:

- Vad är Big Data och vilka tillämpningsområden, nytta, utmaningar samt risker är relaterade till Big Data?

Kortfattat har vi kommit fram till att Big Data definieras utifrån behovet i verksamheten. Olika personer och verksamheter använder olika dimensioner, V_n , för att definiera Big Data. Till skillnad från tidigare teknik som t.ex. data warehousing så öppnar Big Data upp möjligheten att analysera semi- och ostrukturerad data.

Det ger potentiellt ny kunskap genom att hitta mönster som en organisation eller teknisk specialist kan använda för att förändra verksamheten. Analyser som görs i realtid kan ge möjligheten att få ett informationsövertag gentemot konkurrenten. Dock finns risker att informationen kan användas på ett sätt inte som var tänkt. Den personliga integriteten blir begränsad eller försvinner helt. Mönster och analyser kan tillämpas baserat vilken kontext man som tolkare av data/analysen själv väljer. Vi kan även konstatera att Big Data inte bara är en term utan ett perspektiv att se på data med en konceptuell modell på hur ohanterliga datamängder skall behandlas. Studiens bidrag till ämnesområdet informatik är ökad kunskap om fenomenet Big Data. Vidare har vi även kommit fram till att Big Data kan användas som ett nytt komplement av datakällor i beslutsstödssystem.

6.1. Förslag till fortsatt forskning

Synsättet på Big Data skulle kunna utvecklas genom att undersöka möjligheten att bygga en separerad informationsmodell fri från det tekniska arvet. Att data blir oberoende av maskinen. Då datamängder kontinuerligt kommer att öka så kommer även tekniken att hänga med. Snabbare och lagringsmässigt större hårdvara skulle kunna skapa möjligheter att bygga informationsmodeller där olika strukturer av data aggregeras i en gemensam analys.

En liknande undersökning skulle kunna göras om några år för att undersöka hur begreppet har förändrats med tiden. Är det fortfarande Big Data man pratar om eller har organisationers fokus och databehov förändrats?

7.Källor

- Bharal, P. & Halfon, A., 2013, Making sense of Big Data in Insurance, ACORD (Association for Cooperative Operations Research and Development) and MarkLOGIC
- Bhatia, Anuradha, 2013, BIG Data – A Review, IJESRT
- Bollier, D., 2010, The Promise and Peril of Big Data, The Aspen Institute, Communications and Society Program
- Boyd, D. & Crawford, K., 2012, Critical Questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon, Information, Communication & Society
- Braun, V. & Clarke, V., 2006, Using thematic analysis in psychology, Qualitative Research in Psychology
- Chen, M., Mao, S., Liu, Y., 2014, Big Data: A Survey, Mobile Networks and Applications
- Ching-Hsien, Hsu, 2014, Intelligent big data processing, Future Generation Computer Systems
- Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C., 2009, MAD Skills: New Analysis Practices for Big Data, VLDB
- Collaborative, 2012, Making sense of Big Data. A Collaborative point of view, Collaborative White Paper Series
- Corrigan, D., Deutsch, T., Giles, J., Parasuraman, K. & Zikopoulos, P., 2012, Harness the Power of Big Data The IBM Big Data Platform, McGraw Hill Professional
- Cumley, R. & Church, P., 2013, Is “Big Data” creepy?, The International journal of Technology law and practice
- Cuzzocrea, A., Song, I-Y., Davis, K.C., 2011, Analytics over Large-Scale Multidimensional Data: The Big Data Revolution!, Proceedings of the DOLAP’11 Conference
- Davenport, T.H., Barth, P., och Bean, R., 2012, How ‘Big Data’ is Different, MIT Sloan Management Review
- Dijcks, J-P, 2012. How to Implement a Big Data System. [online] Tillgänglig på: <<http://www.oracle.com/technetwork/articles/servers-storage-admin/implementing-bigdata-1502704.html>> [Hämtad 2014-04-11].
- Dumbill, E., 2012. What is big data? An introduction to the big data landscape. [online] Tillgänglig på: <<http://strata.oreilly.com/2012/01/what-is-big-data.html>> [Hämtad 2014-04-22].

- Frankel, David, 2012, Big Data and Risk Management, RIMS, Inc.
- Gobble, MaryAnne M., 2013, Big Data: The Next Big Thing in Innovation, Research Technology Management
- Holme, I. M., Solvang, B. K., 1997, Forskningsmetodik: Om kvalitativa och kvantitativa metoder, Studentlitteratur AB, Lund
- Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey, 2012, BUSINESS INTELLIGENCE AND ANALYTICS:FROM BIG DATA TO BIG IMPACT, MIS Quarterly
- Inmon, W.H., 1996, The Data Warehouse and Data Mining, Communication of ACM
- Larman, C., 2009, Applying UML and patterns,
- Madden, S., 2012, From Databases to Big Data, IEEE Internet Computing
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A., 2011, Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute
- McAfee, A., och Brynjolfsson, E., 2012, Big Data: The Management Revolution, Harvard Business Review
- Michael, K. och Miller, K. W., 2013, Big data: new opportunities and new challenges, Computer
- O'Brien, A., 2012, The Impact of Big Data on Government, IDC Government Insights
- Ohlhorst, F., 2013, Big Data Analytics: Turning big data into big money, John Wiley & Sons, Inc.
- Patel, R., Davidson, B., 2011, Forskningsmetodikens grunder: Att planera, genomföra och rapportera en undersökning, Studentlitteratur AB, Lund
- Pflugfelder, Ehren H., 2014, Big Data, Big Questions, Communication Design Quarterly
- Randal E. Bryant, R.E., Katz, R.H., Lazowska, E.D., 2008, Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society,
- Ross, D., 2013, Big Data and social organisations – A state of the art review, Nominet Trust
- Russom, P., 2011, Big data analytics, TDWI Best Practices Report
- Stubbs, E., 2013, Delivering Business Analytics: Practical Guidelines for Best Practice,
- Swan, M., 2013, The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery, Big Data

- The Open Definition, 2014. The Open Definition. [online] Tillgänglig på:
<<http://opendefinition.org/>> [Hämtad 2014-04-17].
- TIEN, James M., 2013, BIG DATA: UNLEASHING INFORMATION, Systems Engineering Society of China & Springer-Verlag
- Turban, E., Sharda, R., Delen, D. and King, D., 2011, Business Intelligence: A Managerial approach, Pearson
- Villars, R., L. , Eastwood, M., Olofson, C.W., 2011, Big Data: What It IS and Why You Should Care, IDC Analyze the Future
- Zaslavsky, A., Perera, C., and Georgakopoulos, 2012, Sensing as a Service and Big Data, Proceedings of the International Conference on Advances in Cloud Computing (ACC)
- Zikopoulos, P.C., Eaton, C., deRoos, D., Deutsch. T., Lapis, G., 2012, Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Professional

Bilaga 1 - Intervjumall om Big Data i svenska IT bolag v1.0

Inledning (kortfattat)

- > Berätta lite kort om företaget.
- > Berätta om din arbetsuppgift / hur länge du jobbat med detta.
- > Är det okej att vi spelar in denna intervju?
- > Kan vi nämna ditt namn i uppsatsen?

Vad är Big Data enligt yrkesverksamma inom IT området i Sverige?

- > Hur skulle du definiera begreppet *Big Data*?
- > Används det andra ord för begreppet?
- > Vad startade Big Data satsningen hos er?
- > Hur stor del av er verksamhet är big data?
- > Hur mycket arbetar du med Big Data?
- > Har du kommit i kontakt med Big Data innan er satsning?

Vad kan Big Data bidra med till IT bolag verksamma i Sverige?

- > Används Big Data i ert beslutstödssystem (BI), hur?
- > Vill företag ha BI (bättre beslutsstöd, driva förändringar, införa strategier)?
- > Varför räcker det inte med BI, vad kan Big Data ytterligare medföra?
- > På vilket sätt anser du att Big Data hänger ihop med BI området?
- > Vilken kunskap har ni fått från Big Data?
- > Hur ser er Big Data lösning ut ur ett
 - Tekniskt perspektiv?
 - Organisatoriskt?
 - Resursmässigt, hur många arbetar med det, kostnad etc.?

> Vilka nya typer av insikter kan ett företag kan få om den egna verksamheten?

Vilka risker uppfattar IT organisationer i Sverige kring Big Data?

> Vilka organisatoriska risker uppfattar ni med big data?

> Beskriv er riskhantering inom organisationen?

> Är avsaknad av kunskap inom Big Data en risk för er?

> Tillverkar, hyr eller köper ni in er Big Data lösning själva?

> Hur ser ni på konkurrens med att använda lösningar som bygger på Big Data?

> Vad finns det för utmaningar med Big Data (ej tekniska)?

> Vad skulle kunna vara fördelar/nackdelar med Big Data användning?

> Vilka typer av företag/branscher bör överväga Big Data användning?

> I vilka fall bör Big Data inte övervägas som analysmetod?

> Vad kan det finnas för fallgropar med att införa lösningar inom Big Data?

Bilaga 2 - Intervjumall om Big Data i svenska IT bolag v1.1

Inledande frågor

- Berätta lite kort om företaget.
- Berätta om din arbetsuppgift
 - Hur länge du jobbat med detta eller liknande?
 - Hur mycket arbetar du med Big Data?
- Är det okej att vi spelar in denna intervju?
- Kan vi nämna ditt namn i uppsatsen?

Definition av och syn på Big Data

- Hur skulle du definiera begreppet Big Data? Hur ser ni på Big Data?
- Används det andra ord för begreppet?
- Hur förhåller det sig till andra områden så som
 - Data warehouse?
 - Business intelligence?
 - Data analytics?
 - Open Data?
- Hur är Big Data kopplat till ”fenomen” så som
 - Internet of Things?
 - Quantified Self?
 - Social media?

Tillämpningsområden

- Vad startade Big Data satsningen hos er?
- Inom vilka tillämpningsområden jobbar ni idag med Big Data i er organisation?
- Vilka tillämpningsområden är prioriterade idag?
- Vilken typ av system/applikationer är relaterade till Big Data, hur?
- Viken typ av data kan inbegripas i begreppet Big Data?
- Hur ser en Big Data lösning ut ur ett
 - Tekniskt perspektiv?
 - Organisatoriskt?
 - Resursmässigt, hur många arbetar med det, kostnad etc.?
- Vilka typer av företag/branscher bör överväga Big Data användning?
- Vilka möjliga tillämpningsområden ser ni i framtiden?

Nytta och effekter

- Vilken typ av nytta/fördelar kan Big Data bidra till?
- Finns det mål kopplade till Big Data?
- Hur värderar ni nyttan av Big Data?
- Vilken kunskap har ni fått från Big Data?
- Vilka nya typer av insikter kan ett företag få om den egna verksamheten?

Utmaningar

- Vilka utmaningar ser ni med Big Data? (m.a.p. de 4 V:na)
 - Volume – Volymen av data
 - Variety – Olika källor/struktureringsgrader av data
 - Velocity – Hastigheten av datagenerering
 - Veracity – Trovärdighet av data

- Vilka andra utmaningar kan ni se?

Risker och barriärer

- Vilka barriärer ser ni för att lyckas få nytta med Big Data? (exempelvis)
 - Kompetensbarriärer
 - Teknologibarriärer
 - Kulturella barriärer
 - Organisatoriska barriärer
- Vilken typ av risker finns med Big Data?
- Hur kan dessa risker hanteras inom organisationen?
- Vad skulle kunna vara nackdelar/fallgropar med Big Data användning?

Respondentens förslag till övriga frågor

- Avslutningsvis ges möjlighet till den intervjuade att ta upp andra aspekter och frågor kring Big Data.